

(43) International Publication Date
23 October 2014 (23.10.2014)(51) International Patent Classification:
H04N 7/14 (2006.01)(21) International Application Number:
PCT/US2014/032784(22) International Filing Date:
3 April 2014 (03.04.2014)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
13/866,989 19 April 2013 (19.04.2013) US(71) Applicant: **AT&T INTELLECTUAL PROPERTY I, L.P.** [US/US]; 675 W. Peachtree Street, Suite 4000, Atlanta, Georgia 30308 (US).(72) Inventors: **YAROSH, Svetlana**; 422 Ogden Ave #3, Jersey City, New Jersey 07307 (US). **KRISHNAN, Shankar**; 26 Long Hill Ln, Chatham, New Jersey 07928 (US).(74) Agent: **TREMENTOZZI, Ralph**; 304 Indian Trace, #750, Weston, Florida 33326 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM,

AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— with international search report (Art. 21(3))

(54) Title: SYSTEM AND METHOD FOR PROVIDING SEPARATE COMMUNICATION ZONES IN A LARGE FORMAT VIDEOCONFERENCE

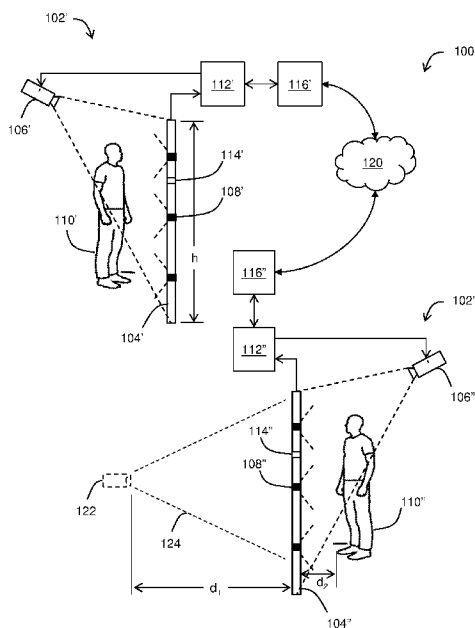


FIG. 1

(57) Abstract: A system that incorporates the subject disclosure performs, for example, displaying a video image of a remote scene at a display surface, wherein the remote scene is remote from the display surface. Overlapping video images are obtained from different vantage points of a local scene observable from the display surface. A composite video image is generated of the local scene from the video images and forwarded to the remote location. A first audio signal is generated representing first sounds associated with a first region of the local scene without representing other sounds associated with a second region of the local scene. The first audio signal is forwarded to audio processing equipment at the remote location to present the first sounds at a first region of the remote scene at the remote location without presenting the first sounds at a second region of the remote scene. Other embodiments are disclosed.

**SYSTEM AND METHOD FOR PROVIDING SEPARATE
COMMUNICATION ZONES IN A LARGE FORMAT VIDEOCONFERENCE**

Inventors

Svetlana Yarosh

Shankar Krishnan

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to U.S. Patent Application Serial Number 13/866,989 filed April 19, 2013. All sections of the aforementioned application is incorporated herein by reference in its entirety.

FIELD OF THE DISCLOSURE

[0002] The subject disclosure relates to a system and method for providing separate communication zones in a large format videoconference.

BACKGROUND

[0003] As the adoption of videoconferencing technology is becoming more widespread in both home and office environments, there are trends toward an immersive experience, referred to as telepresence. Systems supporting a telepresence experience can include telepresence sites having one or more traditional video cameras and relatively large displays, e.g., projection displays that can offer life-sized images of participants at remote telepresence sites. As bandwidth and other infrastructure begin to support larger video displays for videoconferencing, inconsistencies will become more apparent between how a traditional camera usually captures an image for videoconferencing and the metaphor that is most natural for remote interaction. For example, when the various participants are at a range of different distances from a local video camera, captured images of those participants closest to the video camera may be distorted, or even cropped if any part of them happens to extend outside of a field of view of the video camera. Consequently, presentation of the participants on a display wall of a remote telepresence site will

appear distorted and/or cropped. Accordingly, it is not readily apparent where a camera or cameras can be placed, and how each camera's video should be processed to produce a meaningful illusion of presence.

[0004] The "TouchLight" imaging touch screen and display for gesture-based interaction, developed by Microsoft® Research division of Microsoft® Corporation of Redmond, WA, is an example of a display that provides a camera located behind a semi-transparent display screen. Unfortunately, such configurations impose restrictions on the types of displays, as well as on requirements for unobstructed physical space behind the display. It is unclear how such display systems could be mounted on an existing surface, such as a conference room wall, without opening the wall to provide access to physical space located behind the wall.

BRIEF DESCRIPTION OF THE DRAWINGS

[0005] Reference will now be made to the accompanying drawings, which are not necessarily drawn to scale, and wherein:

[0006] FIG. 1 depicts an illustrative embodiment of a telepresence system;

[0007] FIG. 2 depicts an example of a stitched image produced by the illustrative embodiment of the telepresence system of FIG. 1;

[0008] FIG. 3 depicts a top view of an illustrative embodiment of an integrated video camera array and video display surface of a telepresence system;

[0009] FIG. 4 depicts an illustrative embodiment of performance enhancements offered by the telepresence system of FIG. 1;

[00010] FIG. 5A depicts an illustrative example of different proxemic boundaries;

[00011] FIG. 5B depicts a top view of an illustrative embodiment of a telepresence system supporting communications within different proxemic zones;

[00012] FIG. 6 depicts an illustrative embodiment of a telepresence display;

[00013] FIG. 7 depicts an illustrative embodiment of a modular video camera array;

[00014] FIG. 8 depicts a schematic block diagram of an illustrative embodiment of a portion of a telepresence system;

[00015] FIG. 9 depicts a flow diagram of an illustrative embodiment of a process supporting telepresence communications within different proxemic zones operating in portions of the system described in FIGs. 1-8;

[00016] FIGs. 10-11 depict illustrative embodiments of communication systems that provide media services supporting videoconferencing as in FIGs. 1 and 8;

[00017] FIG. 13 depicts an illustrative embodiment of a web portal for interacting with the communication systems to support videoconferencing and/or telepresence as in FIGs. 1 and 8; and

[00018] FIG. 14 is a diagrammatic representation of a machine in the form of a computer system within which a set of instructions, when executed, may cause the machine to perform any one or more of the methods described herein.

DETAILED DESCRIPTION

[00019] The subject disclosure describes illustrative embodiments of providing separate communication zones within a large format videoconference. Other embodiments are included in the subject disclosure.

[00020] One embodiment of the subject disclosure includes a system including a display device to provide a video display of a remote scene at a display surface, wherein the remote scene is remote from the system. The system also includes a number of video cameras positioned across the display surface. The video cameras are aimed at a local scene, proximate to the display surface. Each video camera has a respective field of view to capture a respective portion of the local scene. A first microphone is provided to generate a first audio signal representing first sounds from a first region of the local scene without representing other sounds from a second region of the local scene. A first speaker is provided to generate audio at the first region of the local scene without generating audio at the second region of the local scene. The audio corresponds to sounds originating at a first region of the remote scene. The system also includes a memory to store instructions and a video processor coupled to each video camera. The video processor, responsive to executing the instructions, performs operations including generating a composite image of the local scene from respective portions of the local scene.

[00021] Another embodiment of the subject disclosure includes a process including displaying, by a system comprising a processor, a video image of a remote scene at a display surface. The video image of the remote scene is obtained from video processing equipment at a location that is remote from the system. A number of overlapping video images of a local scene observable from a vantage point of the display surface is obtained, by the system. The overlapping video images are obtained from a number of different vantage points spatially separated across the display surface. A composite video image is generated of the local scene from the overlapping video images. The composite video image is forwarded to the video processing equipment at the remote location. A first audio signal is generated representing first sounds associated with a first region of the local scene without representing other sounds associated with a second region of the local scene. The first audio signal is forwarded to audio processing equipment at the remote location for presentation by the audio processing equipment of the first sounds at a first region of the remote scene at the remote location without presentation by the audio processing equipment of the first sounds to a second region of the remote scene at the remote location.

[00022] Yet another embodiment of the subject disclosure includes a computer-readable storage device comprising instructions that when executed by a processor, cause the processor to perform operations including displaying a video image of a remote scene of a remote location at a display surface. The remote scene is obtained from video processing equipment at a remote location geographically separated from the display surface. A number of overlapping video images is obtained of a local scene observable from a vantage point of the display surface. The multiple overlapping video images are obtained from a number of different vantage points spatially separated across the display surface. A composite video image of the local scene is generated from the overlapping video images. The composite video image is forwarded to the video processing equipment at the remote location. A first audio signal is generated representing first sounds associated with a first region of the local scene without representing other sounds associated with a second region of the local scene. The first audio signal is forwarded to audio processing equipment at the

remote location for presentation by the audio processing equipment of the first sounds at a first region of the remote scene at the remote location without presenting by the audio processing equipment the first sounds to a second region of the remote scene at the remote location.

[00023] In order to support video-mediated communication with large format displays, the problem of camera positioning needs to be solved as it relates to indexicality (e.g., an ability to point or look at something in the remote space and have the remote person be aware of your pointing direction or gaze) and social distance as in proxemics. Proxemics, a subcategory of the study of nonverbal communications, relates to one's use of space as a specialized elaboration of culture. A concept of a personal reaction bubble has been attributed to Edward T. Hall that identifies radially differentiated zones about an individual. Such zones include an intimate space as being a zone extending from an individual to about 1.5 feet away from the individual. So called personal space extends from about 1.5 feet to about 4 feet, social space from about 4 feet to about 12 feet, and public space from about 12 feet to about 25 feet. Thus, proxemics can indicate how far individuals want to be from each other while talking. Intimate conversations might take place in the intimate space, while more general conversations can take place within the personal or social spaces. The distance dictated by proxemics is halved when communicating with others by way of a videoconference or telepresence screen, as the telepresence screen can be perceived as being between the participants. Current video cameras used with large display telepresence systems are not necessarily responsive to such socially appropriate distances, especially for intimate conversations. Participants approaching the display surface to realize conversations in the personal or intimate spaces will be subject to the video distortions and/or cropping of traditional video cameras.

[00024] The techniques disclosed herein include an array of video cameras that can include a one dimensional, e.g., spaced apart along a line, or a multi-dimensional, e.g., spaced apart in a 2-dimensional, array of video cameras (sometimes referred to as a "camera mesh"). Overlapping video images obtained from respective fields of view of the different video cameras of the array are processed into a composite video image and sent to the remote participant. It has been recognized by the authors that two-

dimensional arrays of video cameras can be built into existing display surfaces, such as video display screens and existing environments. Video display screens can include cathode ray tube displays, light-emitting diode (LED) displays, electroluminescent displays, plasma displays, liquid crystal displays (e.g., thin film transistor liquid crystal (LCD) displays, LED backlit LCD displays), organic light emitting diode displays, surface-conduction electron-emitter displays, electronic paper, laser displays, projection displays, such as digital light processing (DLP) displays and the like. Respective video frames of the overlapping video images from the video cameras of the video camera array are combined, or otherwise “stitched” into composite video frames of a composite video image for transmission.

Advantageously, two-dimensional camera arrays can be mounted to virtually any surface, including amorphously shaped surfaces (for example, a “cave” of a video-chat “tent” including one or more of curved or angled display surfaces) and work without significant modifications to the system. The video camera arrays can also be scaled to any size display surface by adding more cameras.

[00025] The techniques disclosed herein are applicable to a home domain, e.g., extending capabilities of services, such as voice over internet protocol (VoIP) services, including Skype® VoIP communications application commercially available from Microsoft® Corporation of Redmond, WA, to an office domain, extending features and capabilities of videoconference facilities, to laboratory domains supporting collaborative activities, and more generally to other public domains, such as schools, libraries, museums, stadiums, sporting venues, etc.

[00026] FIG. 1 depicts an illustrative embodiment of a telepresence system 100 that includes a first telepresence site including a first telepresence terminal 102' and a second telepresence site including a second terminal 102". In some instances, the first telepresence site is referred to as a “local” site, with the second telepresence site being referred to as a “remote” site with respect to the local site. The phrases local and remote are relative and meant to suggest a physical separation between the first and second telepresence sites. In at least some embodiments, the physical separation is a geographic separation in that the local and remote telepresence sites can be located in

different parts of the same building, different buildings of a common complex of buildings, or separate otherwise unrelated buildings and/or complexes.

[00027] The first telepresence terminal 102' includes a first display surface 104' illuminated by a first projector 106'. The first projector 106' projects a video image of a remote scene captured by the second telepresence terminal 102". The first telepresence terminal 102' also includes a first array of video cameras 108' arranged to be aimed to capture overlapping video images of a local scene proximate to the first telepresence terminal 102'. Each video camera of the first array of video cameras 108' has a respective field of view arranged to capture a respective portion of the local scene. For example, each video camera of the first array of video cameras 108' can be aimed at a common point or location identified at or beyond a distance or range with respect to the display surface 104', e.g., 1', 2.5', 5', 10', 20', 50', 100', infinity. In the illustrative example, the local scene includes an area adjacent to the first display surface 104' including a first participant 110' positioned in front of the first display surface a 104'. As will be described in more detail below, the first display surface 104' is relatively large, e.g., approximately as large or larger than the first participant 110', and the first array of video cameras 108' is positioned in or adjacent to the first display surface 104'.

[00028] Each video camera of the first array of video cameras 108' and the projector 106' are coupled to a first video processor 112'. In at least some embodiments, the first telepresence terminal 102' also includes at least one microphone 114'. The first video processor 112' is configured to combine separate video images obtained from each video camera of the first array of video cameras 108' into a composite video image representative of the remote scene. For video images including a series of frames, e.g., thirty frames per second, the composite video image includes a series of composite frames of the remote scene. The first video processor 112' is also configured to format a composite video signal of the composite video image for example encoding the composite video signal according to one or more video standards. Some examples of video standards include digital video using the MPEG-2 video codec, such as the Advanced Television Systems Committee (ATSC) standards, Digital Video Broadcast (DVB) standards, Integrated Services

Digital Broadcasting (ISDB) standards, and the Digital Multimedia Broadcasting (DMB) standards. Other examples of video standards include analog video standards, such as the National Television System Committee (NTSC) standards, the Phase Alternating Line (PAL) standards, etc.

[00029] An output of the first video processor 112' of the first telepresence terminal 102' is input to a first communication processor 116'. The first communication processor 116' reformats or otherwise conditions the composite output video signal received from the first video processor 112' into a first communication signal for transport to a remote destination. The first communication signal is transported across a network 120 to a remote network destination. In the illustrative embodiment, the remote network destination is associated with a remote telepresence site including a second communication processor 116'' associated coupled to the second telepresence terminal 102''.

[00030] The second telepresence terminal 102'' includes a second display surface 104'' illuminated by a second projector 106''. The second projector 106'' projects an image of a local scene of the first telepresence site captured by the first telepresence terminal 102'. The second telepresence terminal 102'' also includes a second array of video cameras 108'' arranged to capture a remote scene of the second telepresence site, captured by the second telepresence terminal 102''. In the illustrative example, the remote scene includes a second participant 110'' positioned in front of the second display surface a 104''.

[00031] Each of the video camera of the second array of video cameras 108'' and the second projector 106'' are coupled to a second video processor 112''. In at least some embodiments, the second telepresence terminal 102'' also includes at least one microphone 114''. The second video processor 112'' is also configured to combine separate video images from each video camera of the second array of video cameras 108'' into a second composite video image representative of the local scene.

[00032] An output of the second video processor 112'' of the second telepresence terminal 102'' is input to a second communication processor 116''. The second communication processor 116'' reformats or otherwise conditions a second composite output video signal of the composite video image from the second video processor

112'' into a second communication signal for transport to a remote destination. The second communication signal is transported across the network 120 to a remote network destination. In the illustrative embodiment, the remote network destination is the first communication processor 116' associated with the first telepresence terminal 102'.

[00033] As shown in more detail at the second telepresence terminal 102'', the second participant 110'' is standing in front of the second display surface 104'' at a distance d_2 measured from the second display surface 104''. Each video camera of the second array of video cameras 108'' has a respective field of view represented by the dashed lines extending away from the second display surface 104''. At relatively close distances, no single video camera of the second array of video cameras 108'' captures an entire image of the second participant 110''. However, the respective overlapping fields of view of several video cameras of the second array of video cameras 108'' collectively include the entire image of the second participant 110''. As will be described further below, the second video processor 112'' combines video images from the video cameras of the second array of video cameras 108'' into a single composite video image. The composite video image is representative of a composite image processed from the individual video images of the video cameras of the second array of video cameras 108''. A plausible composite video image of the remote scene is such that a virtual, non-traditional video camera 122 positioned at a distance d_1 behind a second display surface 104'' would result in a video signal approximating the composite image. A representative virtual field of view 124 is illustrated with respect to the virtual video camera 122.

[00034] In operation, the first participant 110' at the first telepresence terminal 102' observes the composite image of the second participant 110'' at the second telepresence terminal 102''. The first communication processor 116' receives the second communication signal from the second telepresence terminal 102'' and extracts the second composite video signal. The first video processor 112' decodes the second composite video signal, as required, and otherwise renders a composite video image of the second participant 110'' projected by the first projector 106' upon the first display surface 104'. Likewise, the second participant 110'' at the second

telepresence terminal 102'' observes the composite image of the first participant 110'. The second communication processor 116'' receives the first communication signal from the first telepresence terminal 102' and extracts the first composite video signal. The second video processor 112'' decodes the first composite video signal, as required, and otherwise renders a composite video image of the first participant 110' projected by the second projector 106'' upon the second display surface 104''.

[00035] Microphones 114', 114'' at each of the respective telepresence terminals 102', 102'' capture audio at the respective telepresence terminals. The microphones 114', 114'' (generally 114) produce audio signals that can be processed collectively, with the video at the respective video processors 112', 112'' (generally 112) resulting in an audiovisual signal, or separately by way of audio processors (not shown). In either instance, the audio signals are also transported across the network 120 to provide each respective participant with an audiovisual presentation of the scene including the other participant of the videoconference.

[00036] FIG. 2 depicts an example of a composite or stitched image produced by the illustrative embodiment of the telepresence system of FIG. 1. Respective video images of video feeds obtained from each of the video cameras of either of the video camera arrays 108', 108'' (generally 108) can be combined into a composite image 200. In the illustrative example composite image is representative of a scene at a distance d_5 (FIG. 3) from a surface of the camera array 108 including 18 video cameras arranged in a 3X6 rectangular grid. Portions of the scene captured by respective images 202 from each of the video cameras 108 overlap at least partially with portions of the scene captured by respective images 202 of adjacent video cameras of the array of video cameras 108. At the distance d_5 , the extent of overlap 204', 204'' (vertical and horizontal) with adjacent images is minimal. At distances closer to the surface of the camera array will result in less image overlap 204', 204'' (generally 204), with distances less than a threshold distance d_4 (FIG. 3) lacking any overlap and encountering blind spots between images from adjacent cameras. As such, a minimum distance of operation can be established at a distance d_5 slightly beyond the threshold distance d_4 .

[00037] A process of combining the respective video images 202 into the composite video image can be referred to as “stitching” the video feeds together. The process can also be referred to as forming a camera “mesh” in which multiple overlapping video images are combined into a single composite video image. The video images 200, 202 can represent individual images contained in a video signal. The resulting video image includes a stream of such images, e.g., 30 frames per second. In at least some embodiments the video cameras of the video camera array 108 are synchronized such that each video camera provides a respective video frame at about the same time to facilitate processing of the composite video image 200.

[00038] Current approaches used to generate panoramic images from a series of separate images provide robust systems for stitching images taken by rotating a camera around a single point. The same techniques when applied to stitching images obtained from a translational transformation of a camera shot (e.g., obtained from two adjacent cameras on the same plane), present additional algorithmic concerns due to the effect of parallax.

[00039] Parallax relates to a displacement or difference in an apparent position of an object when viewed along two different lines of sight, as measured by an angle or semi-angle of inclination between the two different lines of sight. A parallax interpretation of an object can be performed utilizing respective fields of view of different video cameras. The principles of the parallax effect allows depth perception because each camera is seeing a slightly different view of the scene and objects in the foreground appear more extremely shifted than the objects in the background. The angles/semi-angles of inclination can be determined or otherwise estimated from the different views. When combined with a known spacing or separation between video cameras, an approximate range to an object within the field of view can be estimated. Thus, unlike rotational stitching systems, translational stitching necessarily considers depth of objects in the imaged scene. An example algorithm for producing camera mesh includes capturing a frame from each camera. Each adjacent set of camera images is rectified (e.g., a mathematical transformation is performed on each image to effectively remove small inconsistencies and lens-specific warping introduced by the camera. Each corresponding point in each adjacent set of images is compared using

parallax to provide insight as to depths of objects in the imaged scene. The relevant images at each of the depth layers identified and stitched separately. For example, a view of objects in a background of a scene is stitched separately from objects in a foreground of the imaged scene.

[00040] In at least some embodiments, image processing is applied to video frames captured by the different video cameras of the video camera array. For example, image processing can be applied to identify one or more objects represented in the video frames. Object identification by way of image processing can include various techniques, such as appearance based methods, e.g., edge matching using edge detection, divide-and-conquer search, greyscale matching, gradient matching, histograms of receptive field responses and large modelbases. Other techniques, without limitation, include so-called feature-based methods, such as interpretation trees, hypothesize and test, and so on. Once one or more objects in a scene have been identified, an estimate of a range from the video camera array, e.g., the display surface, to the object can be estimated, e.g., according to the parallax effect.

[00041] Preparation of a composite video frame can combine representations of one or more identified objects in video frames obtained from one or more video cameras of the array. The resulting composite image is generated to represent the favorable attributes disclosed herein. Namely, the composite video image simulates a virtual window into a remote space. Representation of objects in the composite video frame can depend upon a determined or estimated range to the object. For example, an object that is close to the display surface may be represented in video frames of a subset of the video images, whereas objects farther from the display surface may be represented in a larger subset, or even all of the video images of the video camera array. Image stitching of objects at closer ranges can differ from image stitching of objects that are at greater ranges, for example, applying different corrections for warping, etc.

[00042] FIG. 3 depicts a top view of an illustrative embodiment of an integrated video camera array 308 and video display surface 304 of a telepresence system. In the illustrative embodiment, the video camera array 308 includes a 3X6 rectangular array of video cameras 309. A top view of the array 308 illustrates a separation

distance d_3 between adjacent video cameras 309 of the video camera array 308 along one direction, e.g., a horizontal direction. In the illustrative embodiment, the rectangular array of video cameras 308 is arranged along a planar surface coincident with the display surface 304 of a videoconferencing or telepresence terminal. As illustrated, the video cameras 309 are embedded within or slightly behind the display surface 304.

[00043] Each of the video cameras 309 has a respective field of view 311. For example, standard off-the-shelf webcams such as the LifeCam Show®, commercially available from Microsoft Corporation of Redmond, WA, provide a 67° diagonal field of view. At a distance of 1.5 feet away from the camera the camera will have a field of view is approximately 1.79 feet diagonal. Further, in the case of the LifeCam Show® web camera, the field of view as a 4 x 3 aspect ratio. Thus at a social distance of about 1.5 feet (about 0.25 meters when halved in consideration of the display surface) away from the camera the captured image covers an approximate rectangle of 1.43' by 1.07'. Considering the distance of 1.5 feet as a minimum distance d_5 of operation, the number of video cameras 309, as well as a separation between adjacent cameras 309 of the video camera array 308 can be established. In quarter to adequately cover the 4.29' x 3.21' display using the LifeCam Show® cameras 309, will require a 3 x 3 camera array 308. Different arrangements of the video cameras 309 can be determined according to variations of any of the values, such as field of view or minimum operating distance or display size.

[00044] It is understood that in at least some embodiments the surface of the video camera array 308 can be in front of, at, or behind the display surface 304. It is further understood that in at least some embodiments the surface of the video camera array 308 can be non-planar. By way of illustrative example, a videoconferencing cave or tent can provide a display surface along one or more walls or surfaces of the cave or tent. The display surfaces can be angled with respect to each other, curved, continuous, discontinuous, or any suitable combination thereof. In any of such non-planar display configurations, the video camera array can substantially follow the non-planar display surface with a surface of the video camera array 308 being in front of, at, or behind the non-planar display surface.

[00045] FIG. 4 depicts an illustrative embodiment of performance enhancements offered by the telepresence system of FIG. 1, using the display techniques of FIGs. 2 and 3. A first videoconference participant 410' is standing in front of a display surface 404. The display surface 404 is portraying a video image of a second video conference participant 410". The first participant 410' is standing relatively close to the display surface 404, e.g., within arm's length. A traditional videoconference camera 403 positioned at a top central location of the display surface 404 captures a respective image of the first participant 410' illustrated in a second displayed image 420. In the second display the image 420, only a portion of a head and upper body of the first participant is visible. The limited presentation of the first participant 410' in the second displayed image 420 results from the limited field of view of the traditional videoconference camera 403 positioned at the display surface 404.

[00046] A video camera array 408 positioned at display surface 404 captures another respective image of the first participant 410'. The result is illustrated in a third displayed image 422. In the third display the image 422, the entire body of the first participant 410' is visible. The presentation of the first participant 410' in the third displayed image 422 results from the benefits offered by the video camera array 409 and image stitching or camera meshing techniques disclosed herein. In particular, it is possible for the second participant 410" to perceive a direction at which the first participant is pointing. In this instance, the first participant 410' is pointing to an object 424 held by the second participant 410" as displayed in the first image.

[00047] In at least some embodiments, referring again to FIG. 3, the video camera array 308 is combined with one or more audio devices 313. The audio device 313 can include one or more microphones, speakers or any combination thereof. As shown in the illustration, two audio devices 313 provided for the video camera array 308. The two audio devices 313 are separated along the display surface 304. as will be described in more detail below, one or more of the microphones or the speakers of the audio devices 313 can be selected or otherwise designed to provide a video conference or telepresence audio signal associated with a region of the scenes within the vicinity of the audio device 313 without providing such audio to other regions of the scenes subject to the videoconference. Such localized audio can allow for private

or at least semiprivate conversations between videoconference participants positioned within the region of the scene in the vicinity of the localized audio devices 313.

When regional or localized audio is provided, more than one audio channel will be required to support the videoconference.

[00048] FIG. 5A depicts an illustrative example of different proxemic boundaries. Proxemics refers to a social study of the distances between people during different types of personal interactions. Though there are some cultural differences, there are fairly well-established guidelines. According to these guidelines, an intimate region can be identified within about 1.5 feet from an individual 500. An intimate region boundary 502 is shown extending about the individual at this distance. A personal region can be identified between the intimate region 502 and a distance of about 4 feet (1.2 meters) from the individual 500. A personal region boundary 504 is shown extending about the individual 500 at this distance. Likewise, a social region can be identified between the personal region boundary 504 and a distance of about 10 feet (3 meters) from the individual 500. A social region boundary 504 is shown extending about the individual 500 between the personal region boundary 504 and this distance.

[00049] FIG. 5B depicts a top view of an illustrative embodiment of a telepresence system supporting communications within different proxemic zones. At a first, e.g., local, video conference site 501', a first display surface 503' includes a first localized audio device 513a' towards a right hand portion of the first display surface 503' and a second localized audio device 513b' towards a left-hand portion of the first display surface 503'. Each of the first and second localized audio devices 513a, 513b (generally 513) includes a respective microphone to capture sounds from the local video conference site 501' and a respective speaker to provide at the local video conference site 501' with audio representative of sounds originating from a corresponding location of a remote video conference site 501". Depending upon selection and configuration of the localized audio devices, a region of audio coverage can be determined.

[00050] In the illustrative example, the first localized audio device 513a provides audio coverage within a first region 502a. The extent of the first region 502a, for example, can be selected according to any of the proxemic zones, such as the intimate

zone 502a' illustrated. Shown in phantom is a second display surface 503'' at the remote video conference site 501''. The two sites 501', 501'' are geographically separated, but shown here in juxtaposition for convenience, and to reinforce the perception to participants in each of the sites 501', 501''. The second display surface 503'' also includes a first localized audio device 513a'' towards a left-hand portion of the second display surface 503'' and a second localized audio device 513b'' towards the right-hand portion of the second display surface 503''. As shown, the positions of the respective localized audio devices at each of the local and remote videoconferencing sites are in virtual alignment with respect to each other. Consequently, a second participant 510a'' positioned within an intimate zone 502a'' at the remote video conference site 501'' can participate in a private or semiprivate conversation with the first participant 510a' at the local video conference site 501' without disturbing or being otherwise perceptible to other participants 510b, 510c at either the local or remote videoconferencing sites. Namely, sounds within the intimate zone 502a' of the first site 501' reliably captured by a microphone of the first localized audio device 513a', whereas unwanted sounds outside of the intimate zone 502a' are rejected or otherwise not reliably captured. Such unwanted sounds can be captured at audio signal levels many decibels below desired audio, such that they result in signals as in a background murmur.

[00051] Likewise, sounds within the intimate zone 502a'' of the second site 502'', being similarly captured by a microphone of the second localized audio device 513a'' are converted to audio signals, transported to the first site 501' and selectively reproduced as sounds within the corresponding intimate zone 502a'. Amplitudes and other techniques to shape sound, e.g., by controlling one or more of amplitudes and phases of audio signals can be determined by an audio processing system to prevent sounds from traveling much beyond the intimate zone 502a'.

[00052] More than one such intimate videoconferencing zones 502a', 502b' within a large format videoconference system can be established in a like manner. In the illustrative example a second such intimate videoconferencing zone 502b', 502b'' is provided at an opposite end of the display surface 503', 503''. Sounds within the second intimate zone 502b', 502b'' are likewise reliably captured by a microphone of

the second localized audio device 513b' and reproduced by one or more speakers of the audio device 513b'. Alternatively or in addition, other videoconferencing zones or sub-regions can be established within each participating scene of the video conference. In the illustrative example, the third localized audio device 515 provides audio coverage within a third region 504' of the local video conference site. A similar localized audio device provides complimentary coverage within a third region 504'' of the remote video conference site. In the illustrative example the third localized audio devices 515 are arranged or otherwise configured to provide audio coverage within a region 504 corresponding to a personal proxemic zone.

[00053] Any of a number of different microphones or microphone arrays can be used in the localized audio devices 513, 515. Selection of a particular microphone or microphone array can be made according to or otherwise based on an extent and shape of the corresponding video conferencing zone. Likewise, any number of different speakers or speaker arrays can be used in the localized audio devices 513a', 513a'', 513b', 513b'' (generally 513) and 515. Selection of a particular speaker or speaker or a can be made according to the extent and shape of the corresponding video conferencing zone. Although the localized audio devices 513, 515 are shown as being embedded within the respective display surface 503, it is understood that the localized audio devices 513, 515 can be located together with a video camera array or separate from the video camera array, such that the audio portion of the respective videoconferencing zone is achieved, allowing for a realistic exchange of sounds between sites 501', 501'' according to the particular zones. It is also understood that although the localized audio devices 513, 515 are described as including both a microphone and a speaker, other embodiments are possible in which one or more of the microphone in the speaker are separated from each other, such that the audio portion of the respective videoconferencing zone is achieved.

[00054] Through selection and positioning of one or more of the microphones and speakers of the localized audio devices 513, 515 different regions within a common scene subject to the videoconference can be established to support respective conversations between videoconference participants as if the participants were at the same location. In some embodiments, audio coverage can be adapted to cover a so-

called public region, including substantially an entire area in and around the display surface, e.g., including participants 510a', 510b' and 510c.

[00055] FIG. 6 depicts an illustrative embodiment of a telepresence display 600. The telepresence display 600 includes a display surface 604. The display surface 604, for example, can be a wall or a substantial portion of a wall of a videoconferencing facility. (As videoconferencing becomes more widely accepted, it is understood that videoconferencing facilities can include virtually any location such as a home a business, or public place such as an educational institution.) For example, videoconferencing terminals at geographically separated business offices can be arranged to share a wall or substantial portion of a wall within each respective facility. Accordingly one facility can peer through its display wall 604 as if it were looking into a geographically separated business office. Employees in the vicinity of the shared wall will be able to communicate visually and audibly as if the participants were in a common facility, with a restriction that participants at each facility remain on their respective sides of the virtual wall or window between the two facilities. Beneficially, participants are able to approach their respective display surface 604, i.e., the virtual window, within a minimum distance that can be chosen to be within the intimate proxemic zone.

[00056] The example telepresence display 600 includes three video camera sub arrays 608a, 608b, 608c, each including a 3 x 3 array of video cameras 609. Using the example LifeCam Show® video camera discussed above, it would be possible to adequately cover a display area of 12.87' x 3.21'. The display 600 also includes left and right localized audio devices 613a, 613b and a central wide area audio device 615.

[00057] In at the least some embodiments as shown the display 600 can include a visual prompt to notify participants as to the availability and location of localized audio zones. In the illustrative example the display 600 includes a first sign 617a that a left-hand portion of the display surface 604 within the vicinity of the left localized audio device 613a. Likewise the display 600 includes a second sign 617b at the right-hand portion of the display surface 604 within the vicinity of the right localized audio device 613b. Videoconferencing participants can signal each other as to a desirability

of participating in a localized or otherwise private conversation at any time during the course of a video conference/telepresence. Such signaling can occur by video gestures such as a hand wave towards the localized audio device and/or a vocal gesture through other audio devices such as a wide area audio device 615. Each of the audio devices 613a, 613b, 615 can include one or more microphones and/or speakers under the control of audio processing equipment, not shown.

[00058] FIG. 7 depicts an illustrative embodiment of a modular video camera array 700. The video camera array 700 includes a 3 x 3 arrangement of video cameras 702. The video cameras 702 are spaced apart from each other and retained in a fixed relationship with respect to each other by a supporting structure 704. In some embodiments the supporting structure 704 is a simple frame providing frame elements including rigid struts 708 fastened together in a configuration to support the 3 x 3 arrangement of video cameras 702. In some embodiments, the frame includes one or more mounting features 706 to support mounting of the supporting structure 704 to a surface, such as a display surface, e.g., a wall. The mounting features 706 can include one or more of hooks, loops, brackets or other suitable fasteners. In some embodiments, frame elements 708 of the supporting structure 704 can be configurable. For example one or more frame elements 708 of the supporting structure 704 can be folded from a deployed positioned as shown to a compact configuration as might be suitable for storage and for transportation.

[00059] The supporting structure 704 can include cabling as may be required for signaling and four electrical power between each of the video cameras 702 of the video camera array 700. In the illustrative example, such cabling is run within internal conduits of the frame elements of the supporting structure 704, and collected in a common location, such as the cable extension 710, or pigtail, as shown.

[00060] It is understood that in at least some embodiments the video camera array 700 includes one or more audio devices. In the illustrative example the video camera array 700 includes a microphone 712 and a speaker 714. One or more of the microphone 712 and speaker 714 can be mounted to one or more frame elements 708 of the supporting structure 704.

[00061] FIG. 8 depicts a schematic block diagram of an illustrative embodiment of a portion of a telepresence system 800. The telepresence system 800 includes one or more video camera arrays 802', 802''. In the illustrative example a first one by three video camera array 802' is arranged end-to-end with a second one by three video camera array 802''. The result is a 1 x 6 array of video cameras 804a', 804b', 804c', 804a'', 804b'', 804c'' (generally 804). The telepresence system also includes an audio-video processor 810. The audio-video processor 810 includes a first video processor 806' coupled to each of the video cameras 804a', 804b', 804c' of the first video camera array 802'. Likewise, the audio-video processor 810 includes a second video processor 806'' coupled to each of the video cameras 804a'', 804b'', 804c'' of the second video camera array 802''. Each of the video processors 806', 806'' stitches together or otherwise combines respective video images into a composite video image from each of the video cameras 804 of the respective video camera arrays 802', 802''.

[00062] In some embodiments the audio-video processor 810 also includes a first video encoder 808' coupled to the first video processor 806' and a second video encoder 808'' coupled to the second video processor 806'' as shown. Each of the video encoders 808', 808'' provides encoding to the composite video images into a respective encoded video signal. Examples of video encoding include MPEG encoding schemes. It is understood that in at least some embodiments encoding can include application of security for example by way of encryption, scrambling or other suitable obfuscation of the encoded video signal. The encoded video signal is transported to one or more other telepresence sites participating in a videoconference.

[00063] It is understood that the techniques disclosed herein can be applied to support telepresence between more than two sites. It is also understood that the techniques disclosed herein can be applied to any number and arrangement of video cameras 804 to produce one or more stitched video images captured in one or more encoded video signals or streams for network transport between one or more other sites. In some embodiments, the encoded video signals from both a first and second video camera arrays 802', 802'' can be combined in a single encoded video signal. This can be accomplished for example by connecting all of the video cameras to a common video processor 806'. A determination as to a number of encoded video

streams necessary is at least in part determined by the resolution of the resulting display image. If the image is displayed according to a standard, such as any of the high-definition standards, e.g., 1080i, 1080p, 720p, then a single encoded video signal would suffice. However, if the displayed image is larger, e.g., a multiple of a standard resolution, then more than one encoded video signals can be used, one signal for each standard resolution portion of the display.

[00064] An illustrative embodiment the telepresence system 800 includes a first microphone 814' and a first speaker 816' at one end of the first video camera array 802', and a second microphone 814'' and a second speaker 816'' at an opposite end of the second video camera array 802''. The first microphone 814' and the first speaker 816' are coupled to a first audio processor 812'. Likewise, the second microphone 814'' and the second speaker 816'' are coupled to a second audio processor 812''. Each of the first and second audio processors 812', 812'' can provide a respective audio interface to an audio communication channel between telepresence sites. The audio channels can be communicated independently as shown or combined for example with one or more of the encoded video signals for a combined encoded audio visual signal. In at least some embodiments, the audio processor 812', 812'' identifies a source and or destination of audio signals as originating from or being directed to corresponding microphones 814', 814'' and speakers 816', 816'' to allow for selective transport and/or processing of such audio signals.

[00065] FIG. 9 depicts a flow diagram of an illustrative embodiment of a process supporting telepresence communications within different proxemic zones operating in portions of the system described in FIGs. 1-8. A video image is displayed on a display surface at 905. The video image can be displayed using any suitable display technology for video images, such as front projection display (FIG. 1), rear projection display, not shown, and other display devices, such as plasma displays, LCD displays, cathode ray tubes and the like. A remote scene is displayed representative of a remote location, such as a similar video conferencing or telepresence facility (e.g., as shown in FIG. 1).

[00066] Overlapping video images are obtained from different vantage points spatially separated across the display surface at 910. In particular, the overlapping

video images are obtained at a minimum distance from the display surface that fall within one half of the proxemics distance for intimate personal communications (e.g., about 0.25 meters). A composite video image of the local scene is generated from the overlapping video images at 915. An example of such a composite image is shown by the image of a first participant shown in the displayed image 422 (FIG. 4). The composite image represents a plausible composite video image of the local scene that is generated from the overlapping video images, wherein the composite image appears to originate from a virtual non-traditional camera positioned behind the local display surface.. An illustration of an example of such a non-traditional virtual camera is represented by the camera 122 of FIG. 1.

[00067] The composite video image is forwarded to the video processing equipment at the remote location at 920. Processing of the video image can be accomplished locally, e.g., before transporting a video image signal to a remote location, for example, using the video processor 112' in FIG. 1. Alternatively or in addition, some or all of the video processing can be offloaded to one or more other processors, such as the telepresence hub 1030 (FIG. 10).

[00068] A first audio signal representing first sounds associated with a first region of the local scene is generated without representing other sounds associated with a second region of the local scene at 925. The first audio signal is forwarded to the remote location for presentation of the first sounds at a corresponding first region of the remote scene at the remote location without presenting the first sounds to a second region of the remote scene at the remote location at 930.

[00069] FIG. 10 depicts an illustrative embodiment of a first communication system 1000 for delivering media content. The communication system 1000 can represent an Internet Protocol Television (IPTV) media system. Communication system 1000 can be overlaid or operably coupled with any of the telepresence or videoconferencing systems disclosed herein, such as in FIG. 1, as another representative embodiment of communication system 1000. A video image of a remote scene is displayed at the display device 1008. Overlapping video images of a local scene are obtained by the video camera array 1009 from different vantage points spatially separated across the display surface. A composite video image is generated,

appearing to originate from a virtual camera positioned behind the display surface. A first audio signal is generated including first sounds associated with a first region of the local scene without including other sounds associated with a second region of the local scene. The first audio signal is forwarded to equipment at the remote location for presentation of the first sounds at a corresponding first region of the remote scene at the remote location without presenting the first sounds to a second region of the remote scene at the remote location. Other embodiments are disclosed.

[00070] The IPTV media system can include a super head-end office (SHO) 1010 with at least one super headend office server (SHS) 1011 which receives media content from satellite and/or terrestrial communication systems. In the present context, media content can represent, for example, audio content, moving image content such as 2D or 3D videos, video games, virtual reality content, still image content and combinations thereof. The SHS server 1011 can forward packets associated with the media content to one or more video head-end servers (VHS) 1014 via a network of video head-end offices (VHO) 1012 according to a multicast communication protocol.

[00071] The VHS 1014 can distribute multimedia broadcast content via an access network 1018 to commercial and/or residential buildings 1002 housing a gateway 1004 (such as a residential or commercial gateway). The access network 1018 can represent a group of digital subscriber line access multiplexers (DSLAMs) located in a central office or a service area interface that provide broadband services over fiber optical links or copper twisted pairs 1019 to buildings 1002. The gateway 1004 can use communication technology to distribute broadcast signals to media processors 1006 such as Set-Top Boxes (STBs) which in turn present broadcast channels to media devices 1008 such as computers or television sets managed in some instances by a media controller 1007 (such as an infrared or RF remote controller). In at least some embodiments, the media devices 1008 includes a telepresence display according to the techniques disclosed herein employing video camera arrays, image stitching and audio partitioning of a videoconference/telepresence scene. An array of video cameras 1009 can be integrated within or disposed along the display surface of the display device 1008.

[00072] The gateway 1004, the media processors 1006 and media devices 1008 can utilize tethered communication technologies (such as coaxial, powerline or phone line wiring) or can operate over a wireless access protocol such as Wireless Fidelity (WiFi), Bluetooth, Zigbee, or other present or next generation local or personal area wireless network technologies. By way of these interfaces, unicast communications can also be invoked between the media processors 1006 and subsystems of the IPTV media system for services such as video-on-demand (VoD), browsing an electronic programming guide (EPG), or other infrastructure services. One or more of the set-top box 1006 or the gateway 1004 can be configured, for example, by including processing software 1060, in at least some embodiments, to prepare camera mesh of images for images independently obtained by video cameras of the array of video cameras 1009.

[00073] A satellite broadcast television system 1029 can be used in the media system of FIG. 10. The satellite broadcast television system can be overlaid, operably coupled with, or replace the IPTV system as another representative embodiment of communication system 1000. In this embodiment, signals transmitted by a satellite 1015 that include media content can be received by a satellite dish receiver 1031 coupled to the building 1002. Modulated signals received by the satellite dish receiver 1031 can be transferred to the media processors 1006 for demodulating, decoding, encoding and/or distributing broadcast channels to the media devices 1008. The media processors 1006 can be equipped with a broadband port to an Internet Service Provider (ISP) network 1032 to enable interactive services such as VoD and EPG as described above.

[00074] In yet another embodiment, an analog or digital cable broadcast distribution system such as cable TV system 1033 can be overlaid, operably coupled with, or replace the IPTV system and/or the satellite TV system as another representative embodiment of communication system 1000. In this embodiment, the cable TV system 1033 can also provide Internet, telephony and interactive media services.

[00075] The subject disclosure can apply to other present or next generation over-the-air and/or landline media content services system.

[00076] Some of the network elements of the IPTV media system can be coupled to one or more computing devices 1030, a portion of which can operate as a web server for providing web portal services over the ISP network 1032 to wireline media devices 1008 or wireless communication devices 1016.

[00077] Communication system 1000 can also provide for all or a portion of the computing devices 1030 to function as a video conferencing or telepresence hub (herein referred to as telepresence hub 1030). The telepresence hub 1030 can use computing and communication technology to perform function 1062, which can include coordinating and otherwise supporting establishment of a videoconference or telepresence session. Alternatively or in addition, one or more aspects of the image stitching can be offloaded to the telepresence hub 1030, in which multiple video signals are sent from the video camera array 1009 to the telepresence hub 1030. The media processors 1006 and wireless communication devices 1016 can be provisioned with software functions videoconference or telepresence client applications 1062 to utilize the services of the telepresence hub 1030.

[00078] Multiple forms of media services can be offered to media devices over landline technologies such as those described above. Additionally, media services can be offered to media devices by way of a wireless access base station 1017 operating according to common wireless access protocols such as Global System for Mobile or GSM, Code Division Multiple Access or CDMA, Time Division Multiple Access or TDMA, Universal Mobile Telecommunications or UMTS, World interoperability for Microwave or WiMAX, Software Defined Radio or SDR, Long Term Evolution or LTE and so on. Other present and next generation wide area wireless access network technologies can be used in one or more embodiments of the subject disclosure.

[00079] FIG. 11 depicts an illustrative embodiment of a communication system 1100 employing an IP Multimedia Subsystem (IMS) network architecture to facilitate the combined services of circuit-switched and packet-switched systems. Communication system 1100 can be overlaid or operably coupled with any of the telepresence systems disclosed herein, such as the telepresence system 100 of FIG. 1, and communication system 1000 as another representative embodiment of

communication system 1000. One or more video images of a remote scene are transported across the network 1100. Overlapping video images of a local scene are obtained from different vantage points spatially separated across the display surface. A composite video image is generated, in whole or in part by one or more of the functions 1062 on the server 1030 or the functions 1172 on the application server 1117. The composite video image appears to originate from a virtual camera positioned behind the display surface. A first audio signal is generated including first sounds associated with a first region of the local scene without including other sounds associated with a second region of the local scene. The first audio signal is forwarded by the communication system 1100 to equipment at the remote location for presentation of the first sounds at a corresponding first region of the remote scene at the remote location without presenting the first sounds to a second region of the remote scene at the remote location.

[00080] Communication system 1100 can comprise a Home Subscriber Server (HSS) 1140, a tElephone NUmber Mapping (ENUM) server 1130 and other network elements of an IMS network 1150. The IMS network 1150 can establish communications between IMS-compliant communication devices (CDs) 1101, 1102, Public Switched Telephone Network (PSTN) CDs 1103, 1105 and combinations thereof by way of a Media Gateway Control Function (MGCF) 1120 coupled to a PSTN network 1160. The MGCF 1120 need not be used when a communication session involves IMS CD to IMS CD communications. A communication session involving at least one PSTN CD may utilize the MGCF 1120.

[00081] IMS CDs 1101, 1102 can register with the IMS network 1150 by contacting a Proxy Call Session Control Function (P-CSCF) which communicates with an interrogating CSCF (I-CSCF), which in turn, communicates with a Serving CSCF (S-CSCF) to register the CDs with the HSS 1140. To initiate a communication session between CDs, an originating IMS CD 1101 can submit a Session Initiation Protocol (SIP INVITE) message to an originating P-CSCF 1104 which communicates with a corresponding originating S-CSCF 1106. The originating S-CSCF 1106 can submit the SIP INVITE message to one or more application servers (ASs) 1117 that can provide a variety of services to IMS subscribers.

[00082] For example, the application servers 1117 can be used to perform originating call feature treatment functions on the calling party number received by the originating S-CSCF 1106 in the SIP INVITE message. Originating treatment functions can include determining whether the calling party number has international calling services, call ID blocking, calling name blocking, 7-digit dialing and/or is requesting special telephony features (e.g., *72 forward calls, *73 cancel call forwarding, *67 for caller ID blocking, and so on). Based on initial filter criteria (iFCs) in a subscriber profile associated with a CD, one or more application servers may be invoked to provide various call originating feature services.

[00083] Additionally, the originating S-CSCF 1106 can submit queries to the ENUM system 1130 to translate an E.164 telephone number in the SIP INVITE message to a SIP Uniform Resource Identifier (URI) if the terminating communication device is IMS-compliant. The SIP URI can be used by an Interrogating CSCF (I-CSCF) 1107 to submit a query to the HSS 1140 to identify a terminating S-CSCF 1114 associated with a terminating IMS CD such as reference 1102. Once identified, the I-CSCF 1107 can submit the SIP INVITE message to the terminating S-CSCF 1114. The terminating S-CSCF 1114 can then identify a terminating P-CSCF 1116 associated with the terminating CD 1102. The P-CSCF 1116 may then signal the CD 1102 to establish Voice over Internet Protocol (VoIP) communication services, thereby enabling the calling and called parties to engage in voice and/or data communications. Based on the iFCs in the subscriber profile, one or more application servers may be invoked to provide various call terminating feature services, such as call forwarding, do not disturb, music tones, simultaneous ringing, sequential ringing, etc.

[00084] In some instances the aforementioned communication process is symmetrical. Accordingly, the terms “originating” and “terminating” in FIG. 11 may be interchangeable. It is further noted that communication system 1100 can be adapted to support video conferencing. In addition, communication system 1100 can be adapted to provide the IMS CDs 1101, 1102 with the multimedia and Internet services of communication system 1000 of FIG. 10.

[00085] If the terminating communication device is instead a PSTN CD such as CD 1103 or CD 1105 (in instances where the cellular phone only supports circuit-switched voice communications), the ENUM system 1130 can respond with an unsuccessful address resolution which can cause the originating S-CSCF 1106 to forward the call to the MGCF 1120 via a Breakout Gateway Control Function (BGCF) 1119. The MGCF 1120 can then initiate the call to the terminating PSTN CD over the PSTN network 1160 to enable the calling and called parties to engage in voice and/or data communications.

[00086] It is further appreciated that the CDs of FIG. 11 can operate as wireline or wireless devices. For example, the CDs of FIG. 11 can be communicatively coupled to a cellular base station 1121, a femtocell, a WiFi router, a Digital Enhanced Cordless Telecommunications (DECT) base unit, or another suitable wireless access unit to establish communications with the IMS network 1150 of FIG. 11. The cellular access base station 1121 can operate according to common wireless access protocols such as GSM, CDMA, TDMA, UMTS, WiMax, SDR, LTE, and so on. Other present and next generation wireless network technologies can be used by one or more embodiments of the subject disclosure. Accordingly, multiple wireline and wireless communication technologies can be used by the CDs of FIG. 11.

[00087] Cellular phones supporting LTE can support packet-switched voice and packet-switched data communications and thus may operate as IMS-compliant mobile devices. In this embodiment, the cellular base station 1121 may communicate directly with the IMS network 1150 as shown by the arrow connecting the cellular base station 1121 and the P-CSCF 1116.

[00088] It is further understood that alternative forms of a CSCF can operate in a device, system, component, or other form of centralized or distributed hardware and/or software. Indeed, a respective CSCF may be embodied as a respective CSCF system having one or more computers or servers, either centralized or distributed, where each computer or server may be configured to perform or provide, in whole or in part, any method, step, or functionality described herein in accordance with a respective CSCF. Likewise, other functions, servers and computers described herein, including but not limited to, the HSS, the ENUM server, the BGCF, and the MGCF,

can be embodied in a respective system having one or more computers or servers, either centralized or distributed, where each computer or server may be configured to perform or provide, in whole or in part, any method, step, or functionality described herein in accordance with a respective function, server, or computer.

[00089] The telepresence hub 1030 of FIG. 10 can be operably coupled to the second communication system 1100 for purposes similar to those described above. The telepresence hub 1030 can perform function 1062 and thereby provide videoconferencing and/or telepresence services to the CDs 1101, 1102, 1103 and 1105 of FIG. 11. CDs 1101, 1102, 1103 and 1105, which can be adapted with software to perform videoconferencing and/or telepresence function 1170 to utilize the services of the telepresence hub 1030. The telepresence hub 1030 can be an integral part of the application server(s) 1117 performing function 1172, which can be substantially similar to function 1060 and adapted to the operations of the IMS network 1150.

[00090] For illustration purposes only, the terms S-CSCF, P-CSCF, I-CSCF, and so on, can be server devices, but may be referred to in the subject disclosure without the word “server.” It is also understood that any form of a CSCF server can operate in a device, system, component, or other form of centralized or distributed hardware and software. It is further noted that these terms and other terms such as DIAMETER commands are terms can include features, methodologies, and/or fields that may be described in whole or in part by standards bodies such as ^{3rd} Generation Partnership Project (3GPP). It is further noted that some or all embodiments of the subject disclosure may in whole or in part modify, supplement, or otherwise supersede final or proposed standards published and promulgated by 3GPP.

[00091] FIG. 12 depicts an illustrative embodiment of a web portal 1202 which can be hosted by server applications operating from the computing devices 1030 of the communication system 1000 illustrated in FIG. 10. Communication system 1000 can be overlaid or operably coupled with system 1100, communication 1000 and/or communication system 1100 as another representative embodiment of any of the videoconferencing and/or telepresence systems disclosed herein, such as the telepresence system 100 of FIG. 1, communication 1000 and/or communication

system 1100. The web portal can be used to access one or more of the functions 1160, 1162 and 1172, for example, in generating composite video images, audio signals, and the like. The web portal 1202 can be used for managing services of communication systems 1000-1100. A web page of the web portal 1202 can be accessed by a Uniform Resource Locator (URL) with an Internet browser such as Microsoft's Internet Explorer™, Mozilla's Firefox™, Apple's Safari™, or Google's Chrome™ using an Internet-capable communication device such as those described in FIGs. 10-11. The web portal 1202 can be configured, for example, to access a media processor 1006 and services managed thereby such as a Digital Video Recorder (DVR), a Video on Demand (VoD) catalog, an Electronic Programming Guide (EPG), or a personal catalog (such as personal videos, pictures, audio recordings, etc.) stored at the media processor 1006. The web portal 1202 can also be used for provisioning IMS services described earlier, provisioning Internet services, provisioning cellular phone services, and so on.

[00092] The web portal 1202 can further be utilized to manage and provision software applications 1060-1062, and 1172 to adapt these applications as may be desired by subscribers and service providers of communication systems 1000-1100.

[00093] FIG. 13 depicts an illustrative embodiment of a communication device 1300. Communication device 1300 can serve in whole or in part as an illustrative embodiment of the devices depicted in FIGs. 1, and 10-11. The communication device can be used to transport one or more video signals and audio signals between participants of a videoconference or telepresence session.

[00094] To enable these features of supporting videoconferencing and/or telepresence services, communication device 1300 can comprise a wireline and/or wireless transceiver 1302 (herein transceiver 1302), a user interface (UI) 1304, a power supply 1314, a location receiver 1316, a motion sensor 1318, an orientation sensor 1320, and a controller 1306 for managing operations thereof. The transceiver 1302 can support short-range or long-range wireless access technologies such as Bluetooth, ZigBee, WiFi, DECT, or cellular communication technologies, just to mention a few. Cellular technologies can include, for example, CDMA-1X, UMTS/HSDPA, GSM/GPRS, TDMA/EDGE, EV/DO, WiMAX, SDR, LTE, as well

as other next generation wireless communication technologies as they arise. The transceiver 1302 can also be adapted to support circuit-switched wireline access technologies (such as PSTN), packet-switched wireline access technologies (such as TCP/IP, VoIP, etc.) and combinations thereof.

[00095] The UI 1304 can include a depressible or touch-sensitive keypad 1308 with a navigation mechanism such as a roller ball, a joystick, a mouse, or a navigation disk for manipulating operations of the communication device 1300. The keypad 1308 can be an integral part of a housing assembly of the communication device 1300 or an independent device operably coupled thereto by a tethered wireline interface (such as a USB cable) or a wireless interface supporting for example Bluetooth. The keypad 708 can represent a numeric keypad commonly used by phones, and/or a QWERTY keypad with alphanumeric keys. The UI 704 can further include a display 1310 such as monochrome or color LCD (Liquid Crystal Display), OLED (Organic Light Emitting Diode) or other suitable display technology for conveying images to an end user of the communication device 1300. In an embodiment where the display 1310 is touch-sensitive, a portion or all of the keypad 1308 can be presented by way of the display 1310 with navigation features.

[00096] The display 1310 can use touch screen technology to also serve as a user interface for detecting user input. As a touch screen display, the communication device 1300 can be adapted to present a user interface with graphical user interface (GUI) elements that can be selected by a user with a touch of a finger. The touch screen display 1310 can be equipped with capacitive, resistive or other forms of sensing technology to detect how much surface area of a user's finger has been placed on a portion of the touch screen display. This sensing information can be used to control the manipulation of the GUI elements or other functions of the user interface. The display 1310 can be an integral part of the housing assembly of the communication device 1300 or an independent device communicatively coupled thereto by a tethered wireline interface (such as a cable) or a wireless interface.

[00097] The UI 1304 can also include an audio system 1312 that utilizes audio technology for conveying low volume audio (such as audio heard in proximity of a human ear) and high volume audio (such as speakerphone for hands free operation).

The audio system 1312 can further include a microphone for receiving audible signals of an end user. The audio system 1312 can also be used for voice recognition applications. The UI 1304 can further include an image sensor 713 such as a charged coupled device (CCD) camera for capturing still or moving images.

[00098] The power supply 1314 can utilize common power management technologies such as replaceable and rechargeable batteries, supply regulation technologies and/or charging system technologies for supplying energy to the components of the communication device 1300 to facilitate long-range or short-range portable applications. Alternatively, or in combination, the charging system can utilize external power sources such as DC power supplied over a physical interface such as a USB port or other suitable tethering technologies.

[00099] The location receiver 1316 can utilize location technology such as a global positioning system (GPS) receiver capable of assisted GPS for identifying a location of the communication device 1300 based on signals generated by a constellation of GPS satellites, which can be used for facilitating location services such as navigation. The motion sensor 1318 can utilize motion sensing technology such as an accelerometer, a gyroscope, or other suitable motion sensing technology to detect motion of the communication device 1300 in three-dimensional space. The orientation sensor 1320 can utilize orientation sensing technology such as a magnetometer to detect the orientation of the communication device 1300 (north, south, west and east, as well as combined orientations in degrees, minutes, or other suitable orientation metrics).

[000100] The communication device 1300 can use the transceiver 1302 to also determine a proximity to a cellular, WiFi, Bluetooth, or other wireless access points by sensing techniques such as utilizing a received signal strength indicator (RSSI) and/or signal time of arrival (TOA) or time of flight (TOF) measurements. The controller 1306 can utilize computing technologies such as a microprocessor, a digital signal processor (DSP), programmable gate arrays, application specific integrated circuits and/or a video processor with associated storage memory such as Flash, ROM, RAM, SRAM, DRAM or other storage technologies for executing computer

instructions, controlling, and processing data supplied by the aforementioned components of the communication device 1300.

[000101] Other components not shown in FIG. 13 can be used in one or more embodiments of the subject disclosure. For instance, the communication device 1300 can include a reset button (not shown). The reset button can be used to reset the controller 1306 of the communication device 1300. In yet another embodiment, the communication device 1300 can also include a factory default setting button positioned, for example, below a small hole in a housing assembly of the communication device 1300 to force the communication device 1300 to re-establish factory settings. In this embodiment, a user can use a protruding object such as a pen or paper clip tip to reach into the hole and depress the default setting button. The communication device 1300 can also include a slot for adding or removing an identity module such as a Subscriber Identity Module (SIM) card. SIM cards can be used for identifying subscriber services, executing programs, storing subscriber data, and so forth.

[000102] The communication device 1300 as described herein can operate with more or less of the circuit components shown in FIG. 13. These variant embodiments can be used in one or more embodiments of the subject disclosure.

[000103] The communication device 1300 can be adapted to perform the functions of the media processor 1006, the media devices 1008, or the portable communication devices 1016 of FIG. 10, as well as the IMS CDs 1101-1102 and PSTN CDs 1103-1105 of FIG. 11. It will be appreciated that the communication device 1300 can also represent other devices that can operate in communication systems 1000-1100 of FIGs. 10-11 such as a gaming console and a media player.

[000104] The communication device 1300 shown in FIG. 13 or portions thereof can serve as a representation of one or more of the devices of system of FIGs. 1 and/or 8, communication system 1000, and communication system 1100. In addition, the controller 1306 can be adapted in various embodiments to perform the functions 1060, 1062 and 1172, respectively.

[000105] Upon reviewing the aforementioned embodiments, it would be evident to an artisan with ordinary skill in the art that said embodiments can be modified,

reduced, or enhanced without departing from the scope of the claims described below. For example, the videoconferencing or telepresence techniques can be applied to remote learning applications in which a scene containing an educator is shared by way of telepresence with one or more remote learner locations. The educator telepresence facility can include more than one display surface, one for each different remote learning site, whereas each remote learner site can include the same display surface corresponding to the educator telepresence facility. Other embodiments can be used in the subject disclosure.

[000106] It should be understood that devices described in the exemplary embodiments can be in communication with each other via various wireless and/or wired methodologies. The methodologies can be links that are described as coupled, connected and so forth, which can include unidirectional and/or bidirectional communication over wireless paths and/or wired paths that utilize one or more of various protocols or methodologies, where the coupling and/or connection can be direct (e.g., no intervening processing device) and/or indirect (e.g., an intermediary processing device such as a router).

[000107] FIG. 14 depicts an exemplary diagrammatic representation of a machine in the form of a computer system 1400 within which a set of instructions, when executed, may cause the machine to perform any one or more of the methods describe above. One or more instances of the machine can operate, for example, as the telepresence hub 1030, media processor 1006, video processors 112', 112'', communication processors 116', 116'', video processors 806', 806'', video encoders 808', 808'', audio video processor 810, audio processors 812', 812'' and other devices of FIGs. 1-13. In some embodiments, the machine may be connected (e.g., using a network 1426) to other machines. In a networked deployment, the machine may operate in the capacity of a server or a client user machine in server-client user network environment, or as a peer machine in a peer-to-peer (or distributed) network environment.

[000108] The machine may comprise a server computer, a client user computer, a personal computer (PC), a tablet PC, a smart phone, a laptop computer, a desktop computer, a control system, a network router, switch or bridge, or any machine

capable of executing a set of instructions (sequential or otherwise) that specify actions to be taken by that machine. It will be understood that a communication device of the subject disclosure includes broadly any electronic device that provides voice, video or data communication. Further, while a single machine is illustrated, the term “machine” shall also be taken to include any collection of machines that individually or jointly execute a set (or multiple sets) of instructions to perform any one or more of the methods discussed herein.

[000109] The computer system 1400 may include a processor (or controller) 1402 (e.g., a central processing unit (CPU), a graphics processing unit (GPU, or both), a main memory 1404 and a static memory 1406, which communicate with each other via a bus 1408. The computer system 1400 may further include a display unit 810 (e.g., a liquid crystal display (LCD), a flat panel, or a solid state display. The computer system 1400 may include an input device 1412 (e.g., a keyboard), a cursor control device 1414 (e.g., a mouse), a disk drive unit 1416, a signal generation device 1418 (e.g., a speaker or remote control) and a network interface device 820. In distributed environments, the embodiments described in the subject disclosure can be adapted to utilize multiple display units 1410 controlled by two or more computer systems 1400. In this configuration, presentations described by the subject disclosure may in part be shown in a first of the display units 1410, while the remaining portion is presented in a second of the display units 1410.

[000110] The disk drive unit 1416 may include a tangible computer-readable storage medium 1422 on which is stored one or more sets of instructions (e.g., software 1424) embodying any one or more of the methods or functions described herein, including those methods illustrated above. The instructions 1424 may also reside, completely or at least partially, within the main memory 1404, the static memory 1406 and/or within the processor 1402 during execution thereof by the computer system 1400. The main memory 1404 and the processor 1402 also may constitute tangible computer-readable storage media.

[000111] Dedicated hardware implementations including, but not limited to, application specific integrated circuits, programmable logic arrays and other hardware devices that can likewise be constructed to implement the methods described herein.

Application specific integrated circuits and programmable logic array can use downloadable instructions for executing state machines and/or circuit configurations to implement embodiments of the subject disclosure. Applications that may include the apparatus and systems of various embodiments broadly include a variety of electronic and computer systems. Some embodiments implement functions in two or more specific interconnected hardware modules or devices with related control and data signals communicated between and through the modules, or as portions of an application-specific integrated circuit. Thus, the example system is applicable to software, firmware and hardware implementations.

[000112] In accordance with various embodiments of the subject disclosure, the operations or methods described herein are intended for operation as software programs or instructions running on or executed by a computer processor or other computing device, and which may include other forms of instructions manifested as a state machine implemented with logic components in an application specific integrated circuit or field programmable array. Furthermore, software implementations (e.g., software programs, instructions, etc.) can include, but not limited to, distributed processing or component/object distributed processing, parallel processing, or virtual machine processing can also be constructed to implement the methods described herein. It is further noted that a computing device such as a processor, a controller, a state machine or other suitable device for executing instructions to perform operations or methods may perform such operations directly or indirectly by way of one or more intermediate devices directed by the computing device.

[000113] While the tangible computer-readable storage medium 1422 is shown in an example embodiment to be a single medium, the term “tangible computer-readable storage medium” should be taken to include a single medium or multiple media (e.g., a centralized or distributed database, and/or associated caches and servers) that store the one or more sets of instructions. The term “tangible computer-readable storage medium” shall also be taken to include any non-transitory medium that is capable of storing or encoding a set of instructions for execution by the machine and that cause the machine to perform any one or more of the methods of the subject disclosure.

[000114] The term “tangible computer-readable storage medium” shall accordingly be taken to include, but not be limited to: solid-state memories such as a memory card or other package that houses one or more read-only (non-volatile) memories, random access memories, or other re-writable (volatile) memories, a magneto-optical or optical medium such as a disk or tape, or other tangible media which can be used to store information. Accordingly, the disclosure is considered to include any one or more of a tangible computer-readable storage medium, as listed herein and including art-recognized equivalents and successor media, in which the software implementations herein are stored.

[000115] Although the present specification describes components and functions implemented in the embodiments with reference to particular standards and protocols, the disclosure is not limited to such standards and protocols. Each of the standards for Internet and other packet switched network transmission (e.g., TCP/IP, UDP/IP, HTML, HTTP) represent examples of the state of the art. Such standards are from time-to-time superseded by faster or more efficient equivalents having essentially the same functions. Wireless standards for device detection (e.g., RFID), short-range communications (e.g., Bluetooth, WiFi, Zigbee), and long-range communications (e.g., WiMAX, GSM, CDMA, LTE) can be used by computer system 1400.

[000116] The illustrations of embodiments described herein are intended to provide a general understanding of the structure of various embodiments, and they are not intended to serve as a complete description of all the elements and features of apparatus and systems that might make use of the structures described herein. Many other embodiments will be apparent to those of skill in the art upon reviewing the above description. Other embodiments may be utilized and derived therefrom, such that structural and logical substitutions and changes may be made without departing from the scope of this disclosure. Figures are also merely representational and may not be drawn to scale. Certain proportions thereof may be exaggerated, while others may be minimized. Accordingly, the specification and drawings are to be regarded in an illustrative rather than a restrictive sense.

[000117] Although specific embodiments have been illustrated and described herein, it should be appreciated that any arrangement calculated to achieve the same purpose

may be substituted for the specific embodiments shown. This disclosure is intended to cover any and all adaptations or variations of various embodiments. Combinations of the above embodiments, and other embodiments not specifically described herein, can be used in the subject disclosure.

[000118] The Abstract of the Disclosure is provided with the understanding that it will not be used to interpret or limit the scope or meaning of the claims. In addition, in the foregoing Detailed Description, it can be seen that various features are grouped together in a single embodiment for the purpose of streamlining the disclosure. This method of disclosure is not to be interpreted as reflecting an intention that the claimed embodiments require more features than are expressly recited in each claim. Rather, as the following claims reflect, inventive subject matter lies in less than all features of a single disclosed embodiment. Thus the following claims are hereby incorporated into the Detailed Description, with each claim standing on its own as a separately claimed subject matter.

CLAIMS

What is claimed is:

1. A system comprising:
 - a display device having a display surface to provide a video display of a remote scene at the display surface, wherein the remote scene is remote from the system;
 - a plurality of video cameras positioned across the display surface and aimed at a local scene proximate to the display surface, wherein each video camera of the plurality of video cameras has a respective field of view to capture a respective portion of the local scene;
 - a first microphone to generate a first audio signal, the first audio signal representing first sounds from a first region of the local scene without representing other sounds from a second region of the local scene;
 - a first speaker to generate audio at the first region of the local scene without generating audio at the second region of the local scene, wherein the audio corresponds to sounds originating from a first region of the remote scene;
 - a memory to store instructions; and
 - a video processor coupled to each video camera of the plurality of video cameras and in communication with the memory, wherein the video processor, responsive to executing the instructions, performs operations comprising generating a composite image of the local scene from respective portions of the local scene.
2. The system of claim 1, wherein the first region of the local scene comprises an area extending away from a reference location along the display surface to a distance of about 0.3 meters from the reference location.
3. The system of claim 2, further comprising a plurality of audio devices positioned across the display surface to provide a plurality of audio zones within different regions of the local scene.

4. The system of claim 3, wherein a first audio zone of the plurality of audio zones is substantially isolated from a second zone of the plurality of audio zones.
5. The system of claim 1, wherein the operations further comprise:
determining, with respect to the display surface, a range of an object within the local scene; and
generating the composite image of the local scene responsive to the range of the object.
6. The system of claim 5, wherein the determining of the range of the object within the local scene comprises performing a parallax interpretation of the object from respective fields of view of different video cameras of the plurality of video cameras.
7. The system of claim 1, further comprising a supporting structure coupled to the plurality of video cameras, wherein the supporting structure is mountable to the display surface and wherein the supporting structure maintains the plurality of video cameras in a fixed arrangement when mounted to the display surface.

8. A method comprising:
- displaying, by a system comprising a processor, a video image of a remote scene at a display surface, wherein the video image of the remote scene is obtained from video processing equipment at a remote location that is remote from the system;
 - obtaining, by the system, a plurality of overlapping video images of a local scene observable from a vantage point of the display surface, wherein the plurality of overlapping video images are obtained from a plurality of different vantage points spatially separated across the display surface;
 - generating, by the system, a composite video image of the local scene from the plurality of overlapping video images;
 - forwarding, by the system, the composite video image to the video processing equipment at the remote location;
 - generating, by the system, a first audio signal representing first sounds associated with a first region of the local scene without representing other sounds associated with a second region of the local scene; and
 - forwarding, by the system, the first audio signal to audio processing equipment at the remote location for presentation by the audio processing equipment of the first sounds at a first region of the remote scene at the remote location without presentation by the audio processing equipment of the first sounds at a second region of the remote scene at the remote location.

9. The method of claim 8, further comprising:
 - receiving, by the system, from the audio processing equipment at the remote location a second audio signal representing second sounds associated with the first region of the remote scene without representing other sounds associated with a second region of the remote scene; and
 - reproducing, by the system, at the first region of the local scene, the second sounds responsive to the second audio signal without presenting the second sounds at the second region of the local scene,
 - wherein vocal communication is supported between a first participant positioned within the first region of the local scene and a second participant positioned within the first region of the remote scene, the first and second sounds being substantially imperceptible at other regions of either the remote or the local scenes.
10. The method of claim 9, wherein the second sounds are reproduced by a speaker.
11. The method of claim 9, wherein the first region of the local scene comprises an area extending away from a reference location along the display surface to a distance of about 0.3 meters from the reference location.
12. The method of claim 8, wherein the first audio signal is generated by a microphone.
13. The method of claim 8, wherein the generating of the composite video image of the local scene comprises combining respective frames of video images of the plurality of overlapping video images.

14. The method of claim 13, further comprising:
 - determining, with respect to the display surface, a range of an object within the local scene, wherein the combining of the respective frames of the video images of the plurality of overlapping video images depends upon the range of the object.
15. A computer-readable storage device comprising instructions that when executed by a processor, cause the processor to perform operations comprising:
 - displaying a video image of a remote scene at a display surface, wherein the remote scene is obtained from video processing equipment at a remote location geographically separated from the display surface;
 - obtaining a plurality of overlapping video images of a local scene observable from a vantage point of the display surface, wherein the plurality of overlapping video images are obtained from a plurality of different vantage points spatially separated across the display surface;
 - generating a composite video image of the local scene from the plurality of overlapping video images;
 - forwarding the composite video image to the video processing equipment at the remote location;
 - generating a first audio signal representing first sounds associated with a first region of the local scene without representing other sounds associated with a second region of the local scene; and
 - forwarding the first audio signal to audio processing equipment at the remote location for presentation by the audio processing equipment of the first sounds at a first region of the remote scene at the remote location without presentation by the audio processing equipment of the first sounds at a second region of the remote scene at the remote location.

16. The computer-readable storage device of claim 15, wherein the operations further comprise:

receiving from the audio processing equipment at the remote location a second audio signal representing second sounds associated with the first region of the remote scene at the remote location without representing other sounds associated with a second region of the remote scene at the remote location; and

reproducing at the first region of the local scene, the second sounds responsive to the second audio signal without presenting the second sounds at the second region of the local scene,

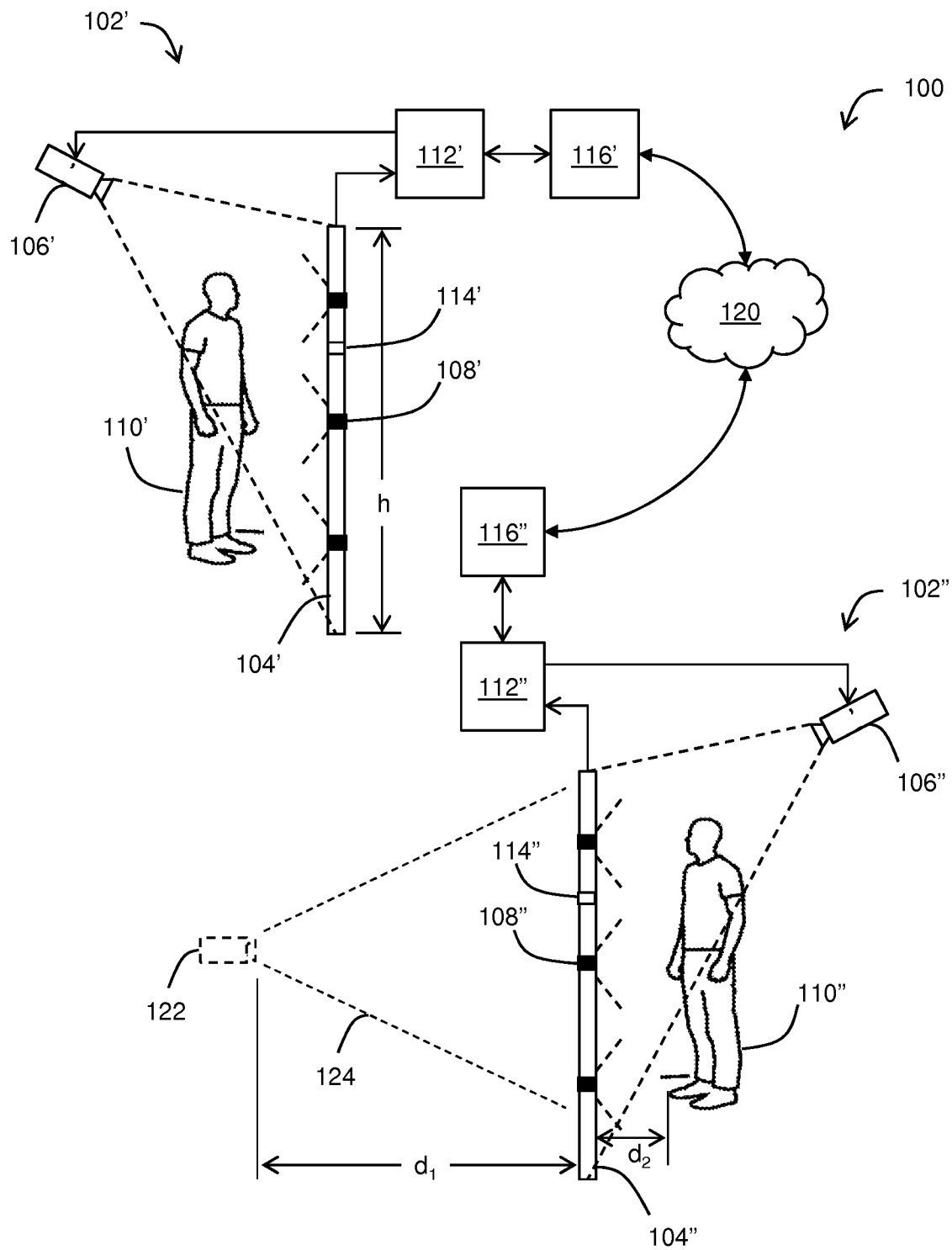
wherein vocal communication is supported between a first participant positioned within the first region of the local scene and a second participant positioned within the first region of the remote scene, the first and second sounds being substantially imperceptible at other regions of either the remote or the local scenes.

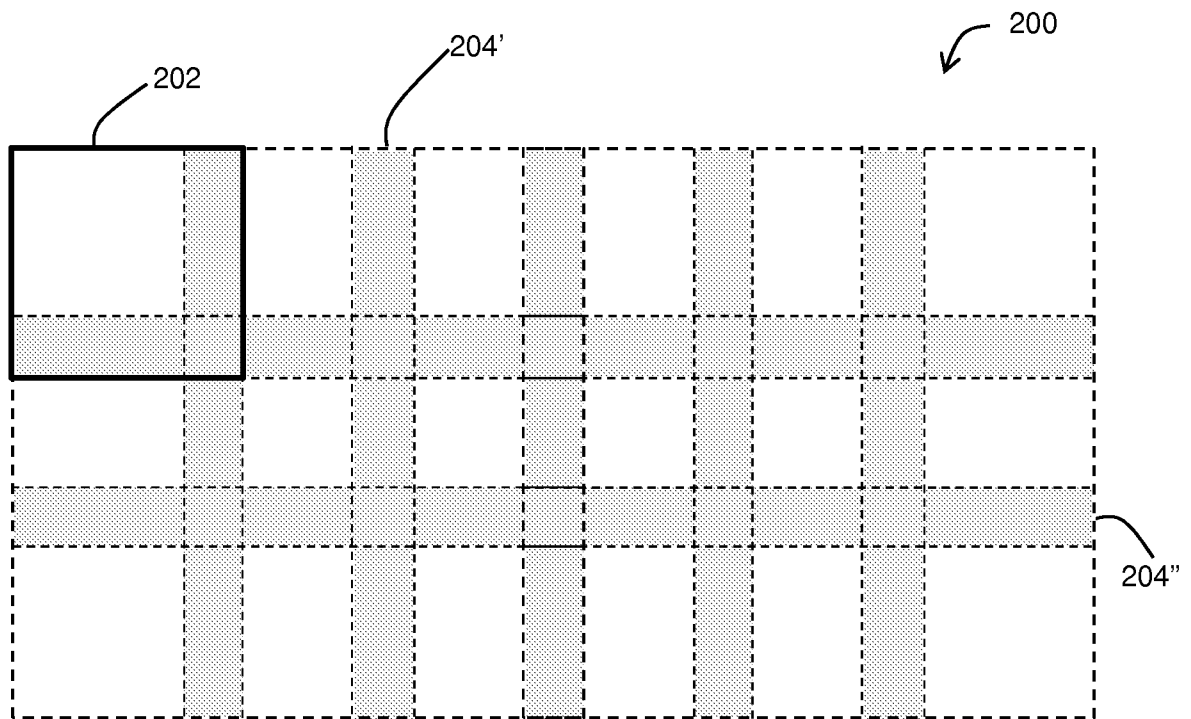
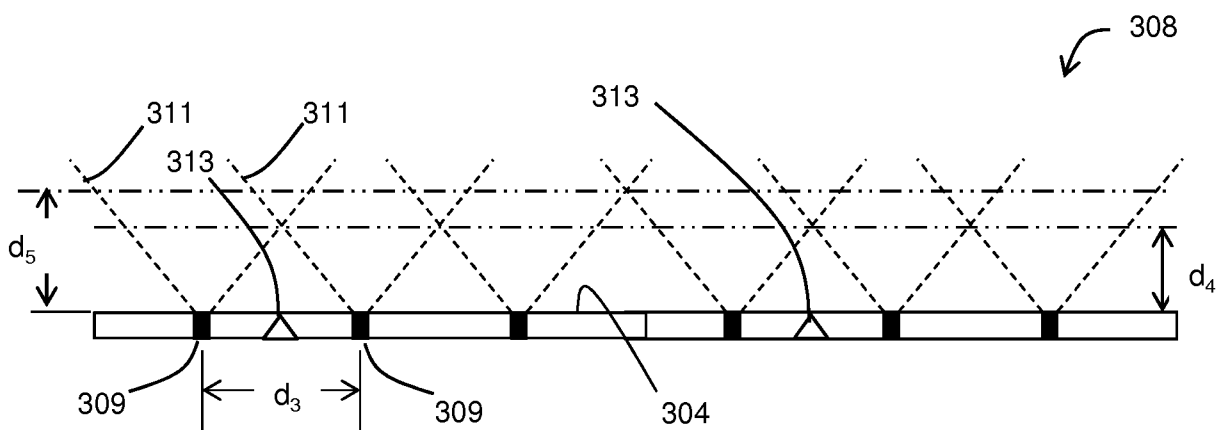
17. The computer-readable storage device of claim 16, wherein the second sounds are reproduced by a speaker.

18. The computer-readable storage device of claim 15, wherein the first audio signal is generated by a microphone.

19. The computer-readable storage device of claim 15, wherein the generating of the composite video image of the local scene comprises combining respective frames of video images of the plurality of overlapping video images.

20. The computer-readable storage device of claim 19, further comprising:
determining, with respect to the display surface, a range of an object within the local scene, wherein the combining of the respective frames of the video images of the plurality of overlapping video images depends upon the range of the object.

**FIG. 1**

**FIG. 2****FIG. 3**

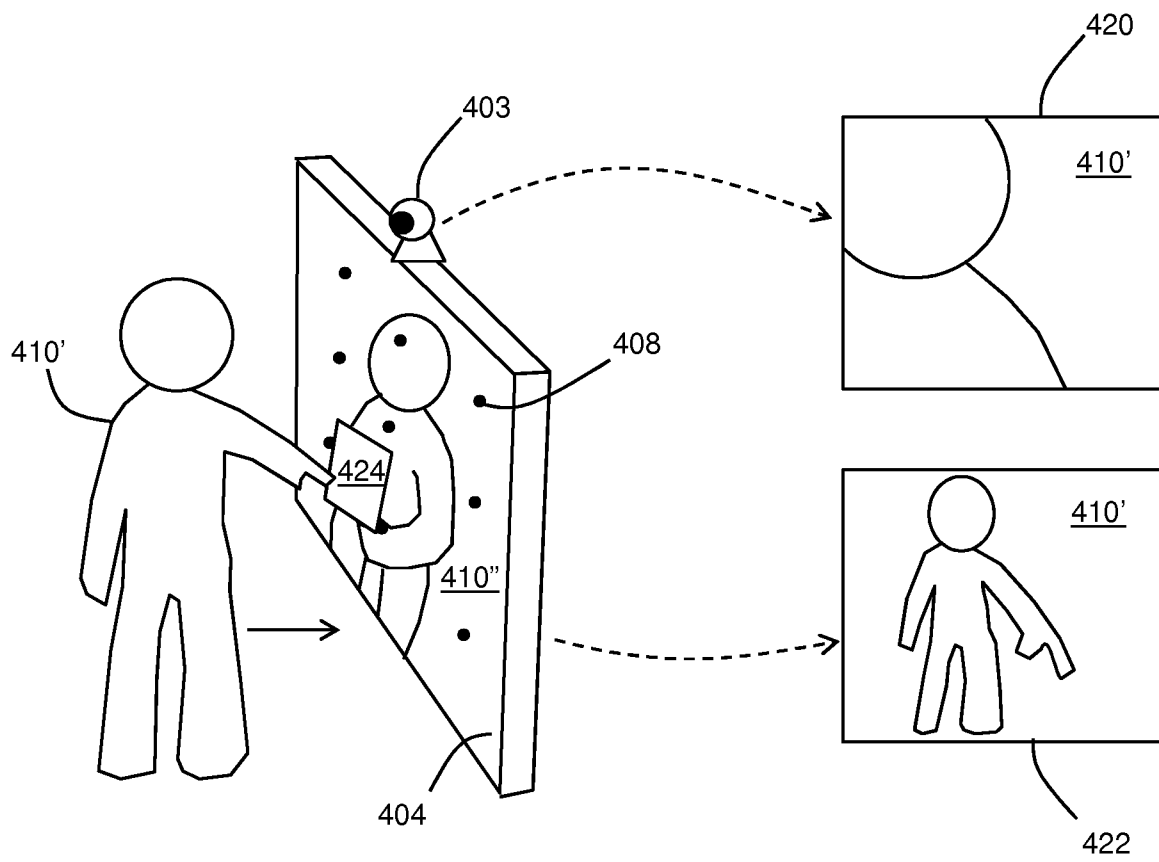


FIG. 4

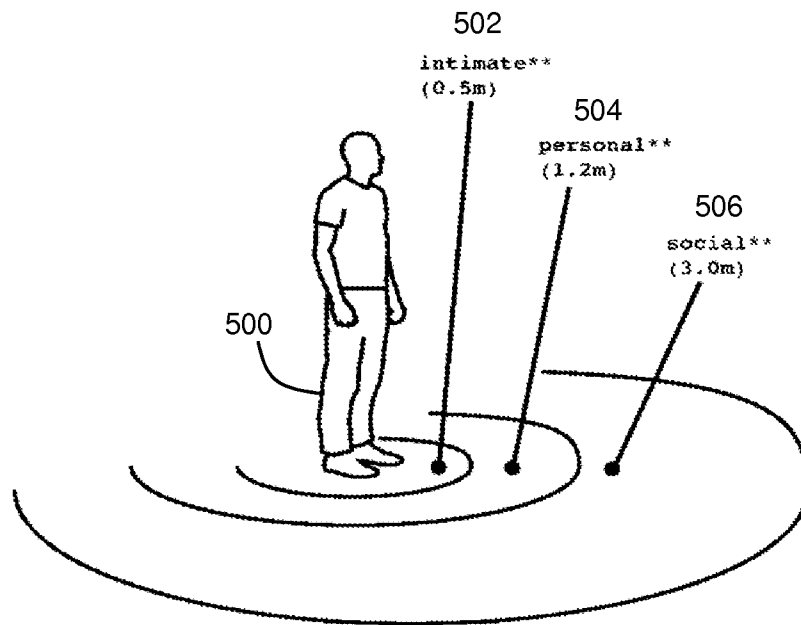


FIG. 5A

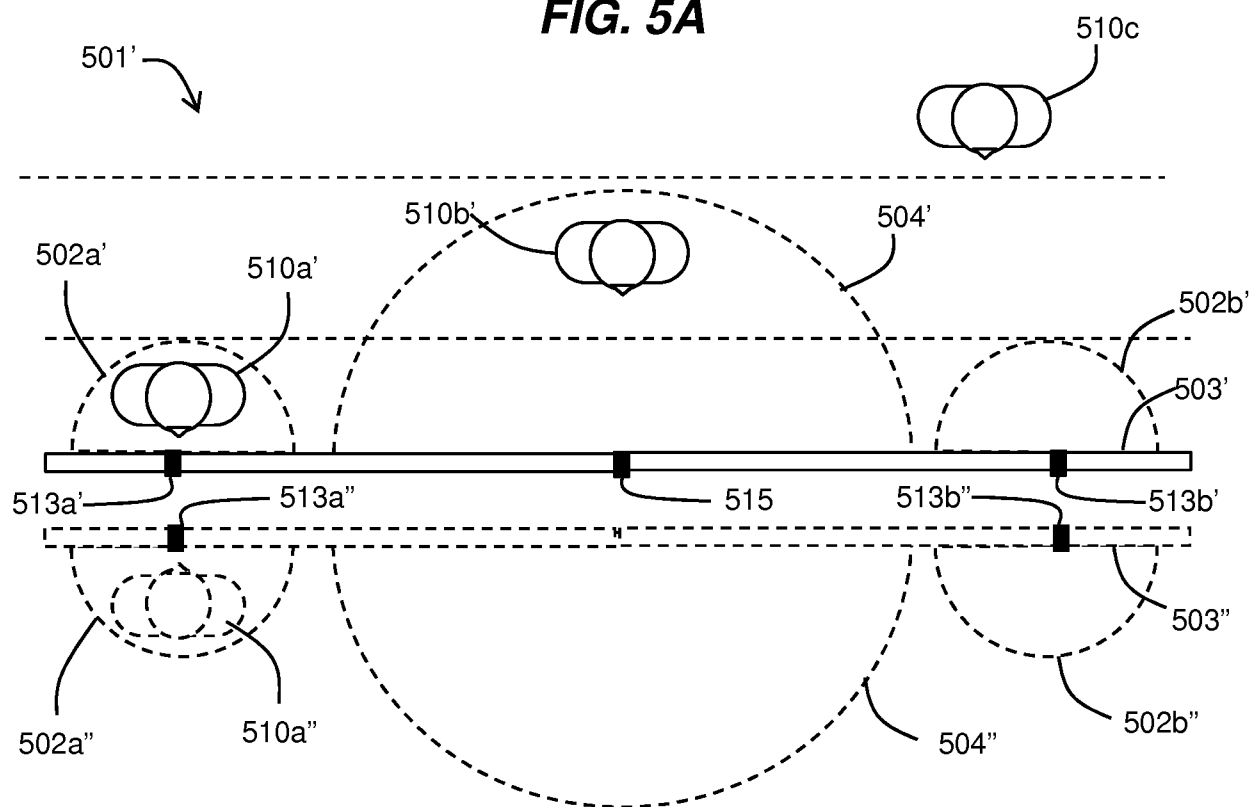


FIG. 5B

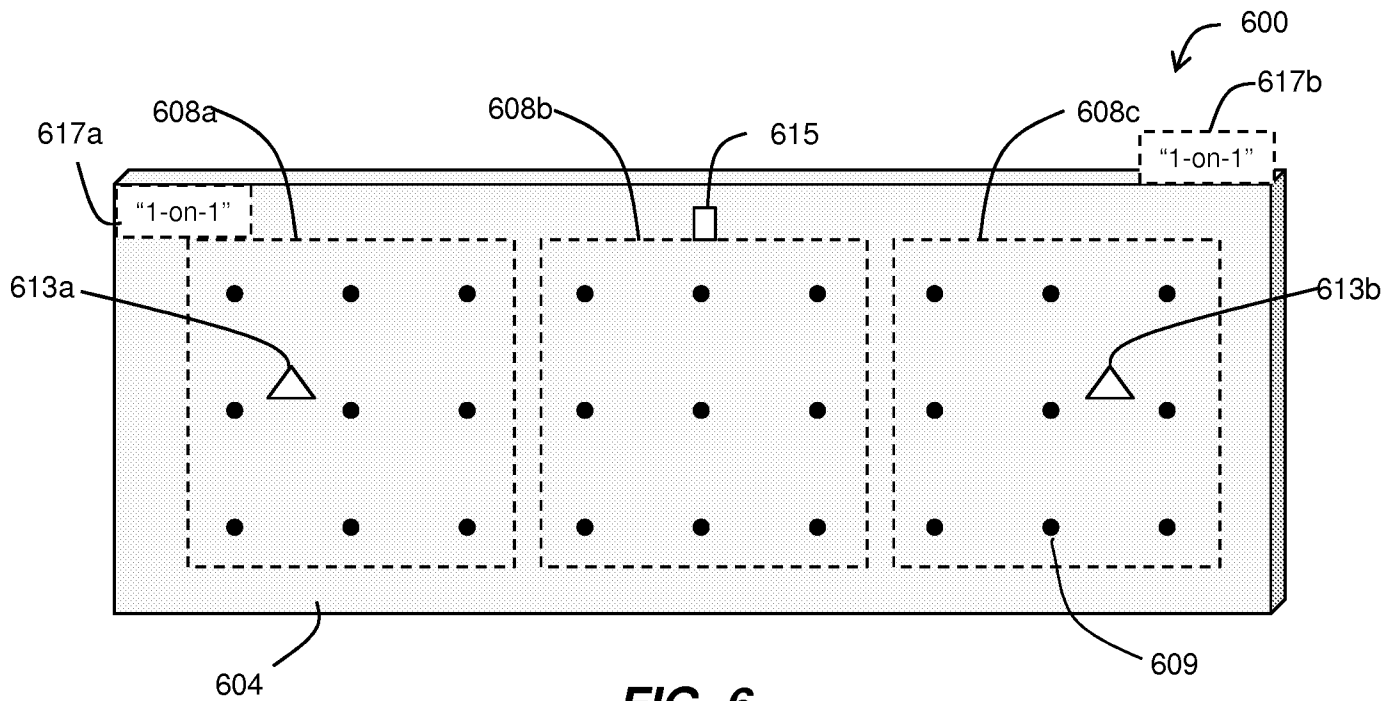


FIG. 6

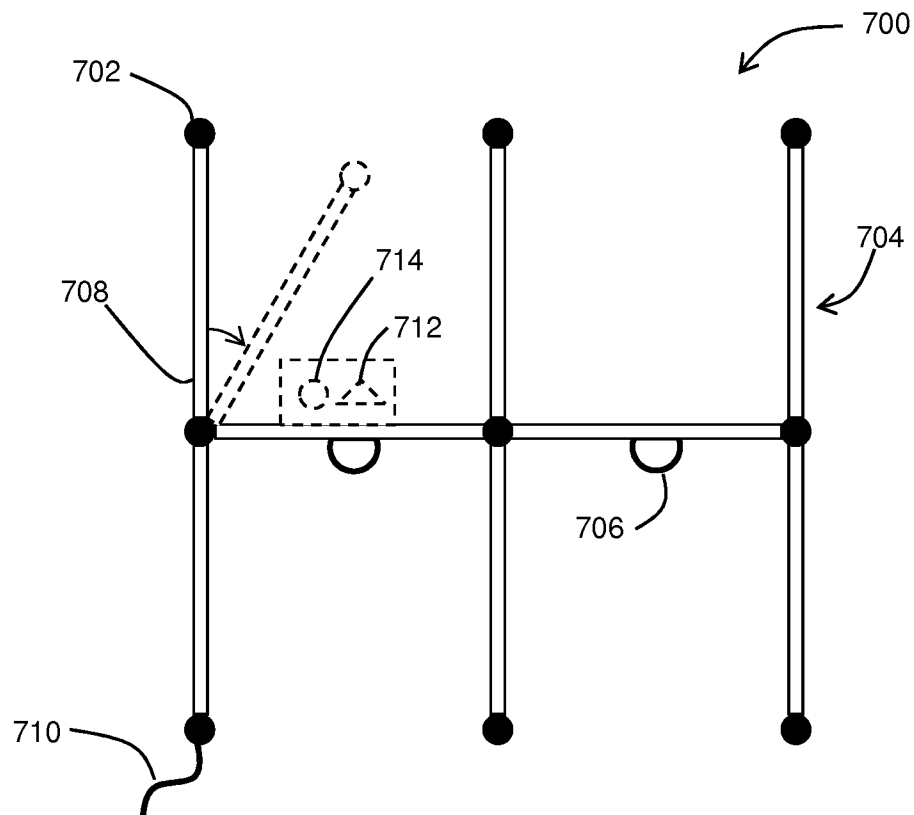
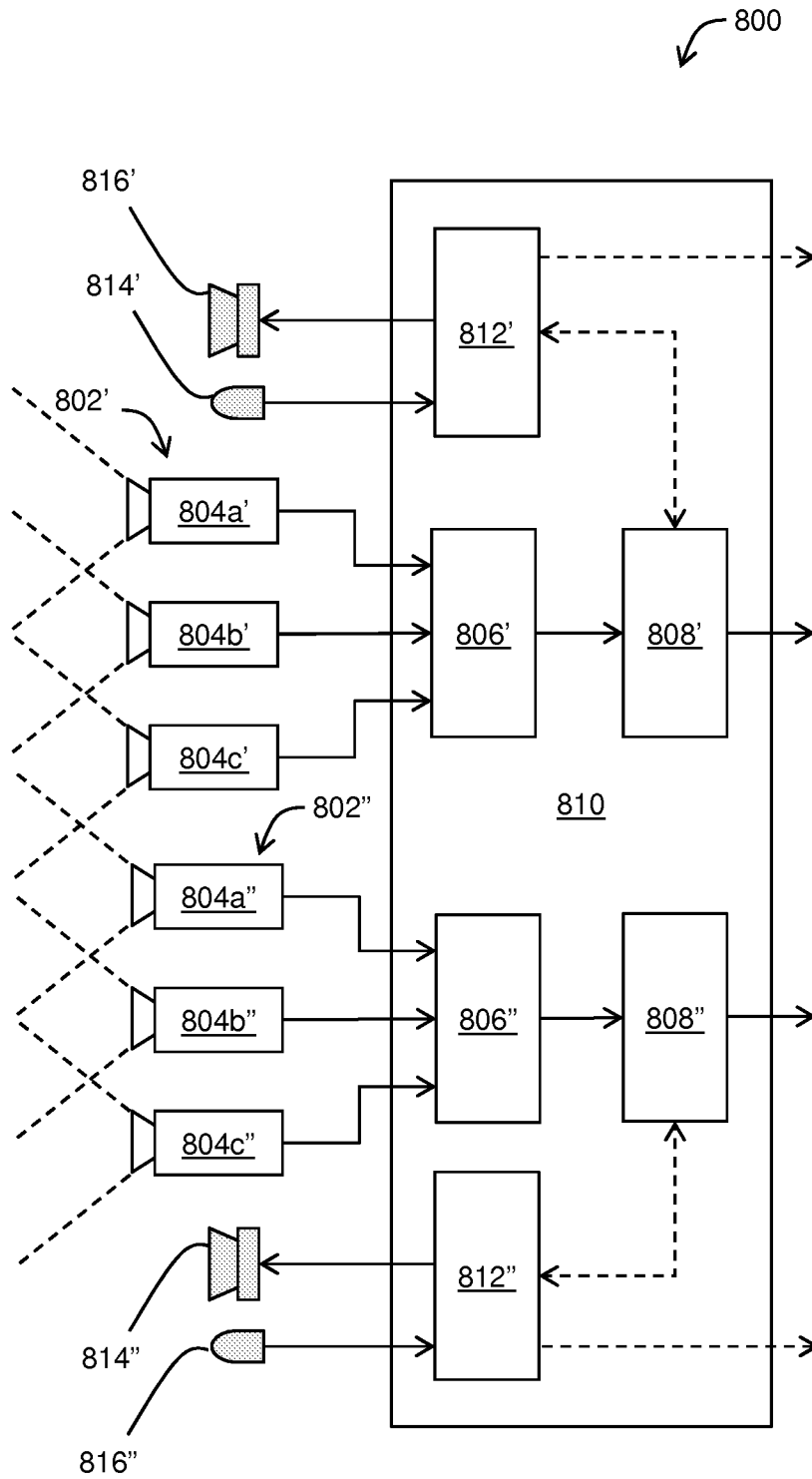
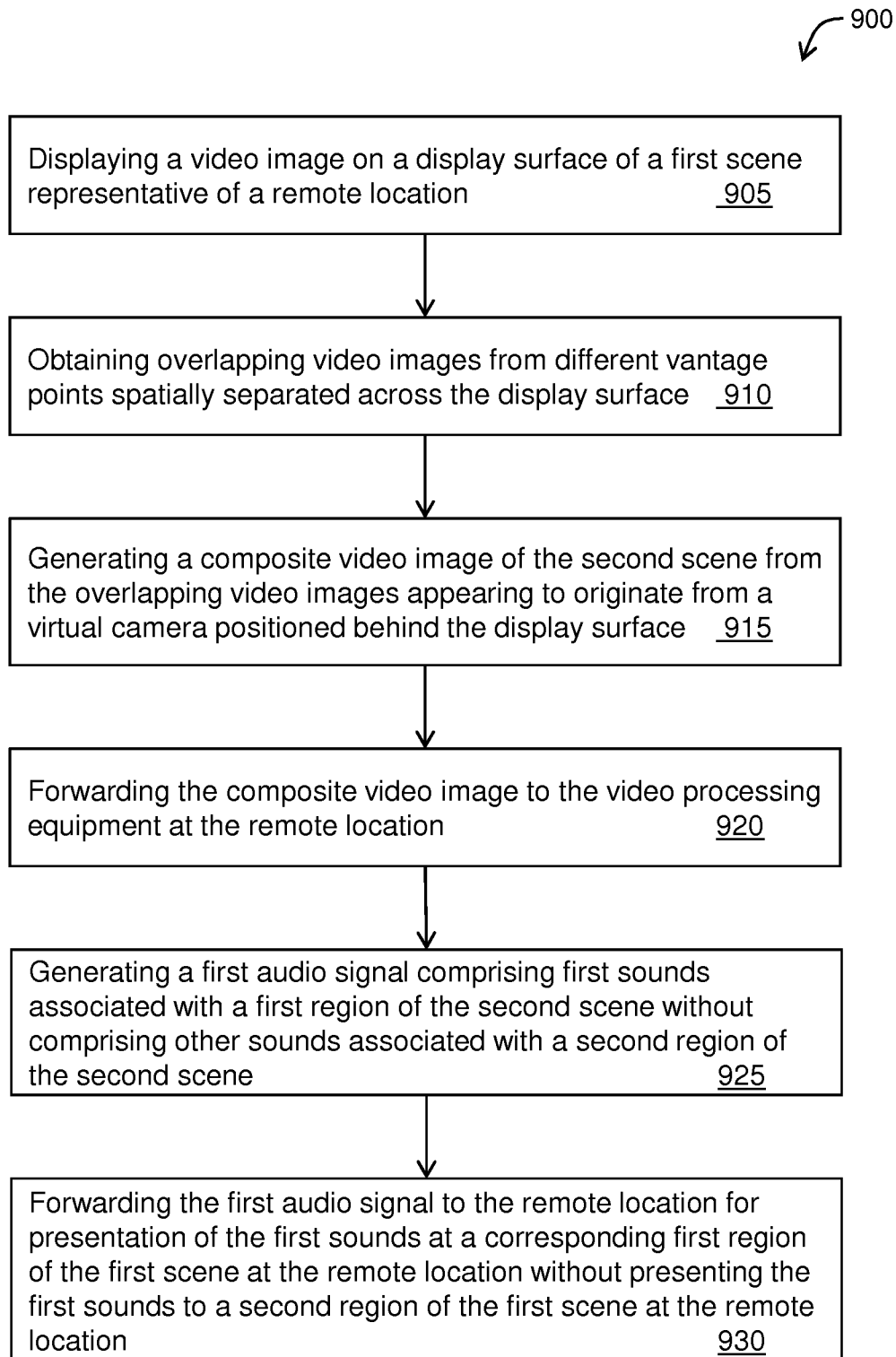


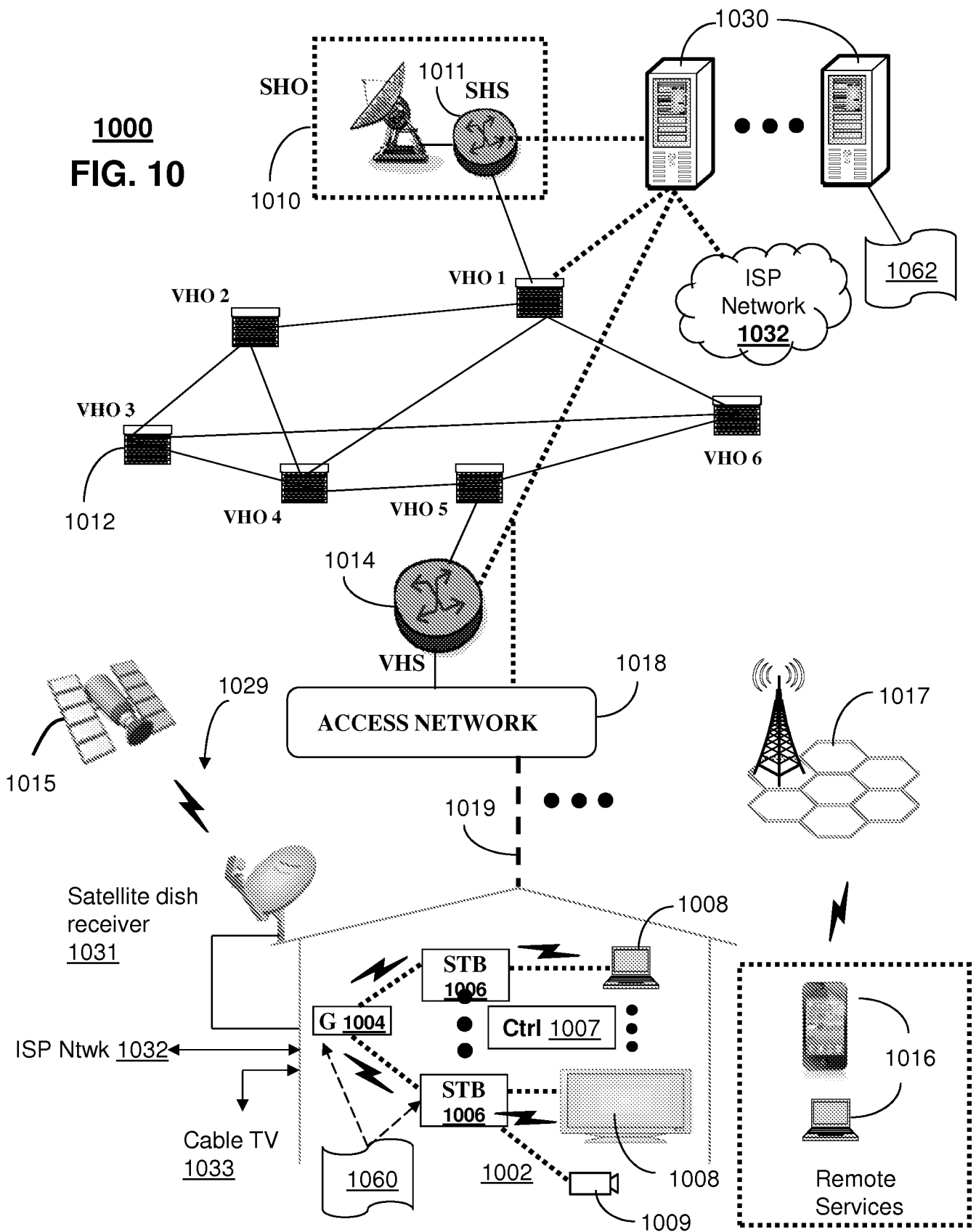
FIG. 7

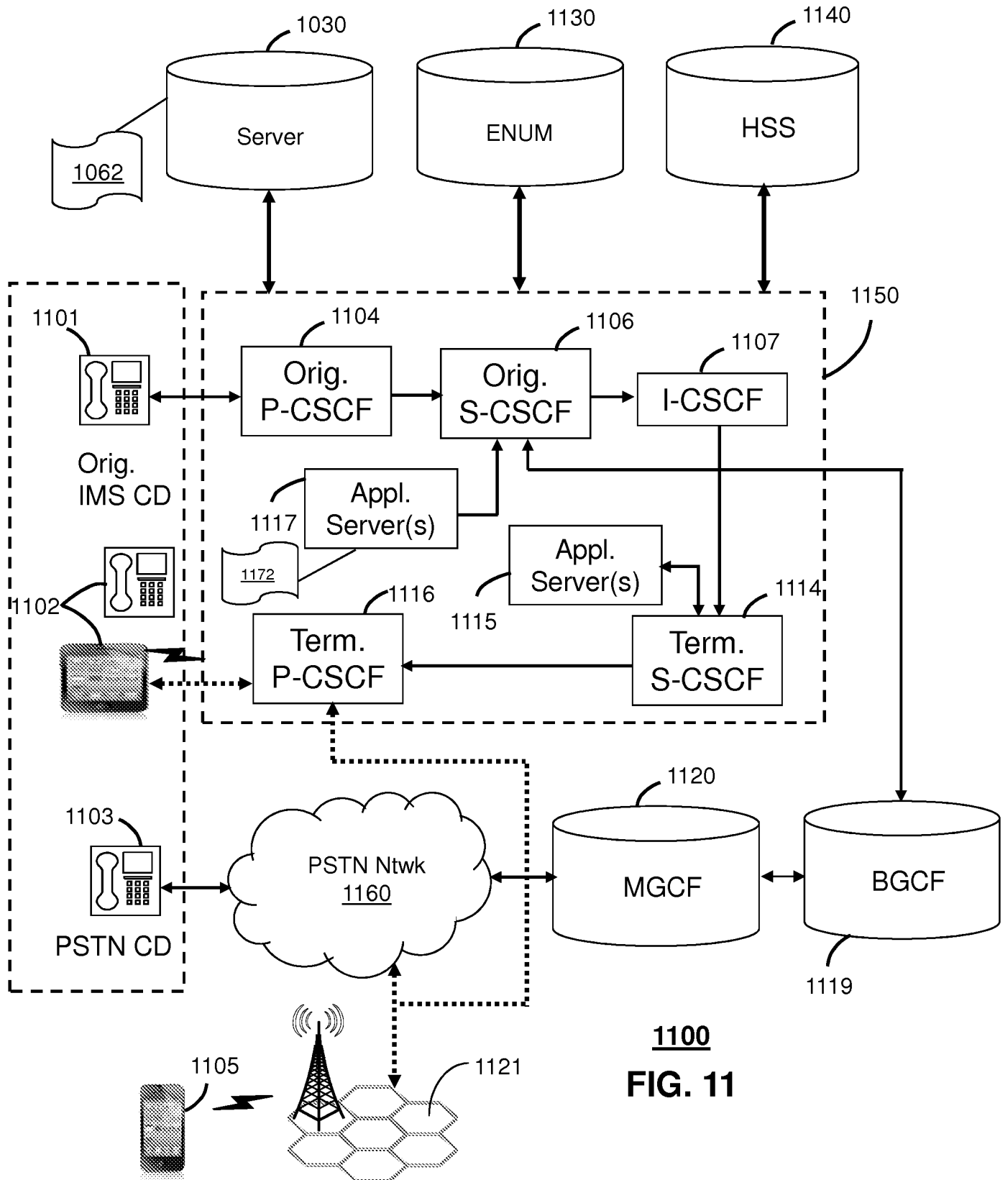
**FIG. 8**

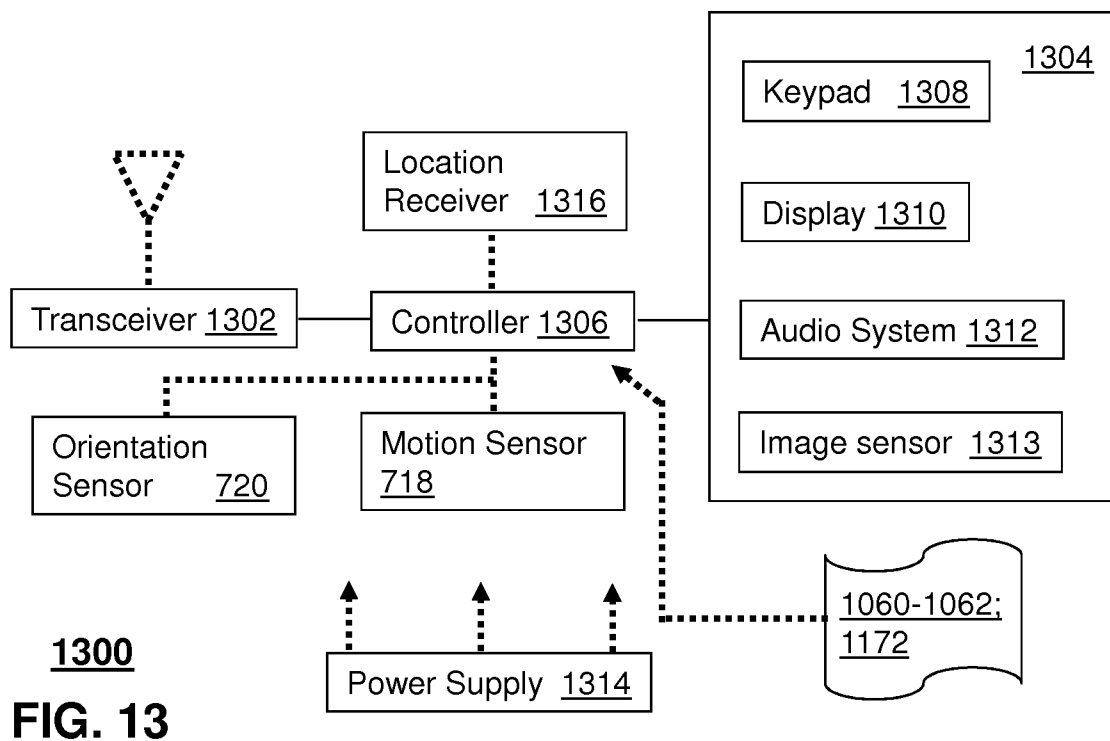
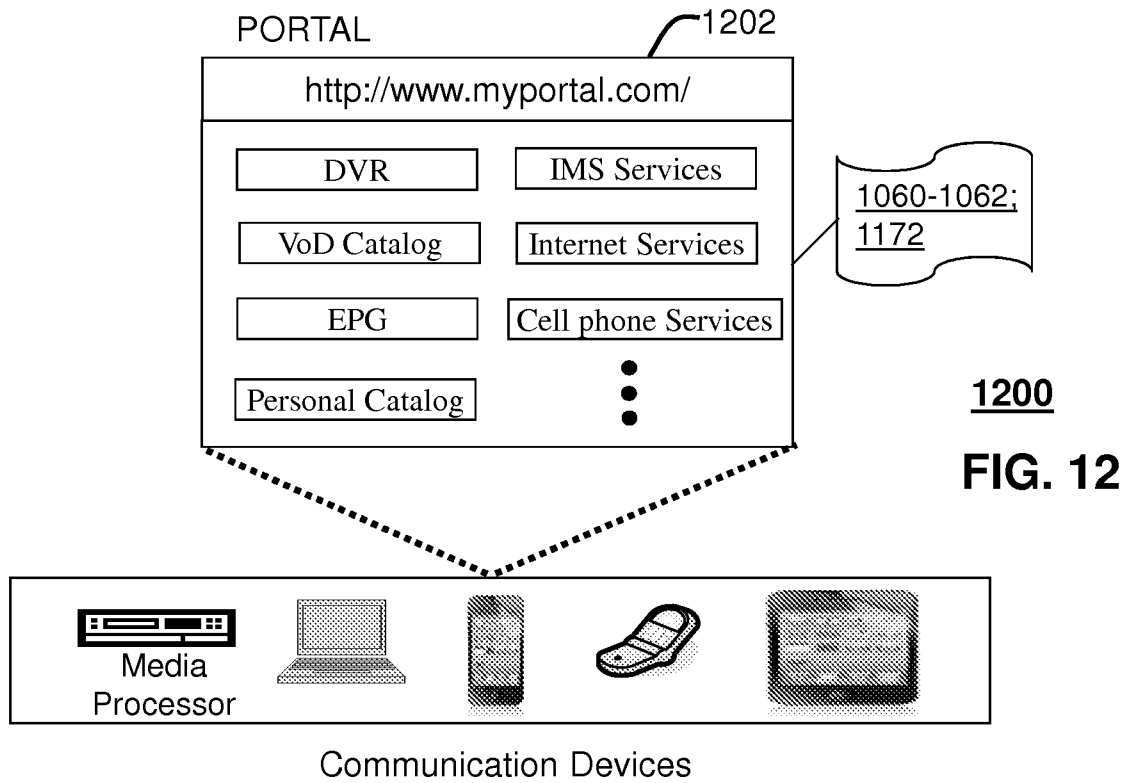
7/11

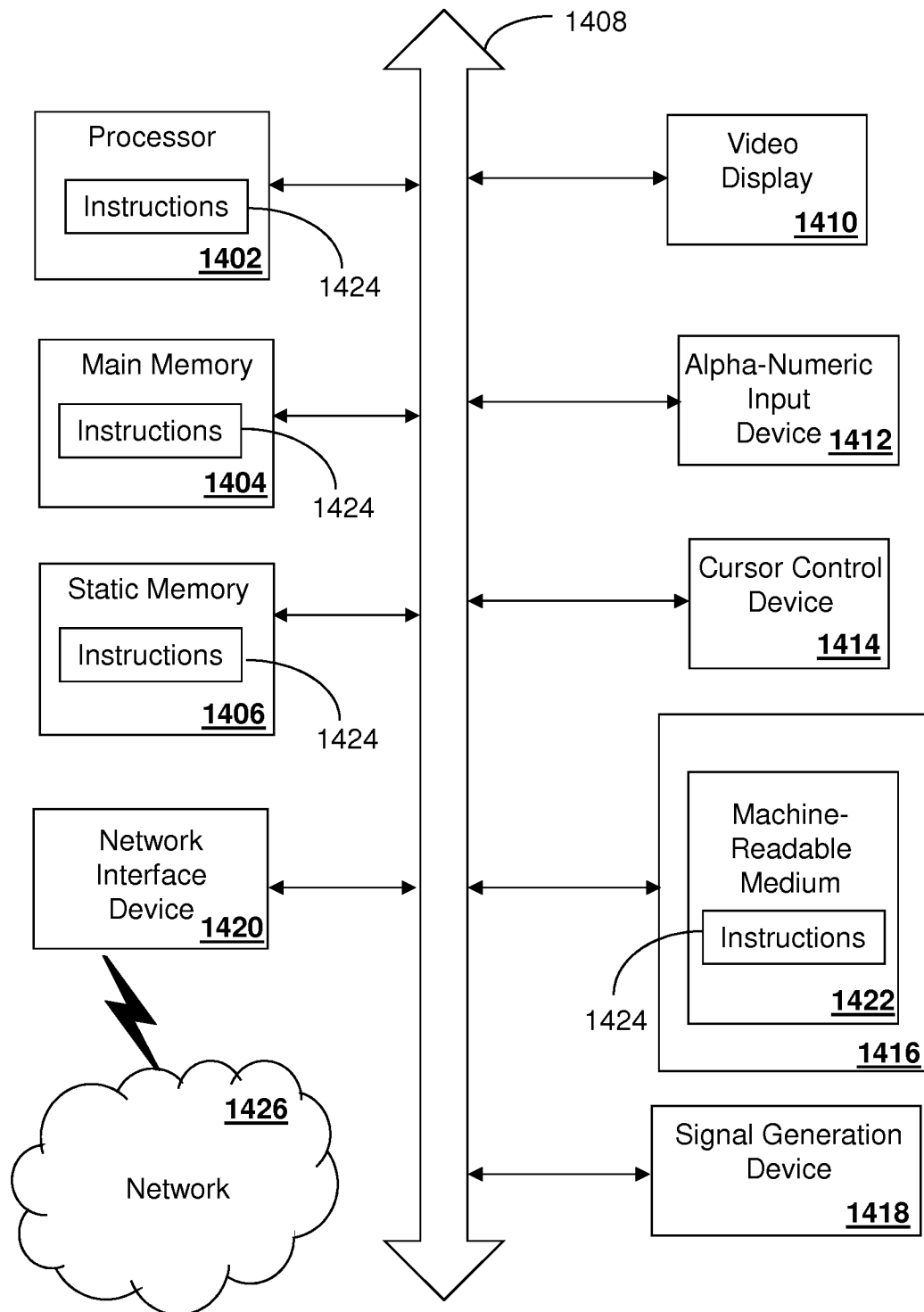
**FIG. 9**

1000
FIG. 10









1400
FIG. 14

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2014/032784

A. CLASSIFICATION OF SUBJECT MATTER
INV. H04N7/14
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
H04N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EP0-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 2010/111275 A1 (CISCO TECH INC [US]; MAUCHLY J WILLIAM [US]) 30 September 2010 (2010-09-30) paragraph [0017] - paragraph [0019] paragraph [0024] paragraph [0029] - paragraph [0039]; figures 6A-6D paragraph [0047]; figure 7 -----	1-20
Y	WO 2007/123946 A2 (CISCO TECH INC [US]; GRAHAM PHILIP R [US]; HARRELL RANDY K [US]; MCLAU) 1 November 2007 (2007-11-01) abstract page 20, line 4 - page 22, line 2; figure 2 ----- -/--	1-20



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

16 July 2014

Date of mailing of the international search report

25/07/2014

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Beaudoin, Olivier

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2014/032784

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 6 624 841 B1 (BUCHNER GEORGES [FR] ET AL) 23 September 2003 (2003-09-23) abstract column 2, line 32 - column 3, line 2; figure 3 -----	1-20

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/US2014/032784

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2010111275 A1	30-09-2010	CN 102362495 A EP 2412161 A1 US 2010245535 A1 WO 2010111275 A1	22-02-2012 01-02-2012 30-09-2010 30-09-2010
WO 2007123946 A2	01-11-2007	EP 2022266 A2 US 2007263079 A1 US 2010214391 A1 WO 2007123946 A2	11-02-2009 15-11-2007 26-08-2010 01-11-2007
US 6624841 B1	23-09-2003	CA 2284884 A1 CN 1257631 A DE 69803168 D1 DE 69803168 T2 DK 0970584 T3 EP 0970584 A1 ES 2171016 T3 FR 2761562 A1 JP 4699574 B2 JP 4885928 B2 JP 2001517395 A JP 2009081861 A US 6624841 B1 WO 9844734 A1	08-10-1998 21-06-2000 21-02-2002 22-08-2002 22-04-2002 12-01-2000 16-08-2002 02-10-1998 15-06-2011 29-02-2012 02-10-2001 16-04-2009 23-09-2003 08-10-1998